

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-167095

(P2001-167095A)

(43) 公開日 平成13年6月22日 (2001.6.22)

(51) Int.Cl.<sup>7</sup>

識別記号

F I

テーマコード(参考)

G 0 6 F 17/30

G 0 6 F 15/40

3 7 0 D 5 B 0 5 0

G 0 6 T 1/00

15/62

P 5 B 0 7 5

7/00

15/70

4 6 0 B 5 L 0 9 6

9 A 0 0 1

審査請求 未請求 請求項の数21 O L (全 29 頁)

(21) 出願番号 特願平11-343256

(22) 出願日 平成11年12月2日 (1999.12.2)

(31) 優先権主張番号 特願平11-272183

(32) 優先日 平成11年9月27日 (1999.9.27)

(33) 優先権主張国 日本 (J P)

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72) 発明者 関口 俊一

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(72) 発明者 井須 芳美

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(74) 代理人 100102439

弁理士 宮田 金雄 (外1名)

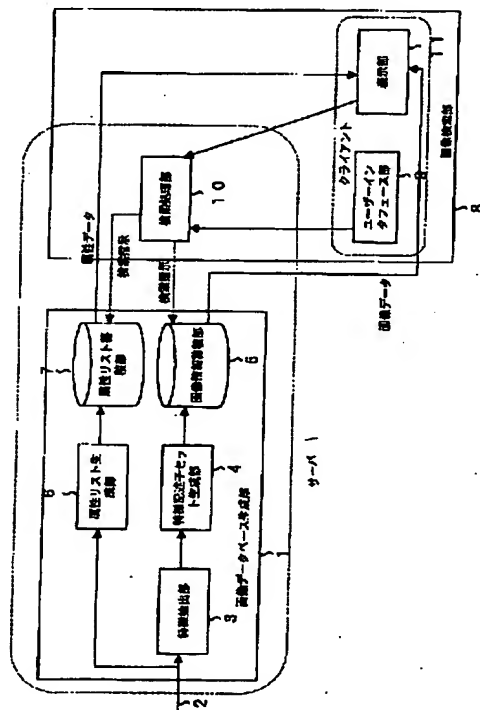
最終頁に続く

(54) 【発明の名称】 画像検索システムおよび画像検索方法

(57) 【要約】

【課題】 画像検索処理を効率良く実行する。

【解決手段】 検索処理部10は、ユーザインターフェース部9より属性リストの検索条件が入力すると、属性リスト蓄積部7に蓄積された属性リストを検索して検索条件に適合した属性情報が表示部11に出力され表示される。次に、検索処理部10は、ユーザインターフェース部9より類似検索の検索条件が入力すると、画像情報蓄積部5に蓄積された画像データを検索して、検索条件に適合する特徴記述子セットを持った画像データを選択して、表示部11に出力し表示させる。



(2)

## 【特許請求の範囲】

【請求項1】 入力画像データから画像特徴量を抽出し特徴記述子を生成する特徴記述子生成部と、生成された特徴記述子を入力画像データと対応づけて蓄積する画像情報蓄積部と、

入力画像データに付随して入力する属性情報に基づいて属性リストを生成する属性リスト生成部と、属性情報に関する検索条件が入力すると前記属性リストを検索して当該検索条件に適合する属性情報を出力すると共に、特徴記述子に関する検索条件が入力すると前記画像情報蓄積部を検索して当該検索条件に適合する画像データを出力する画像検索部と、を備えることを特徴とする画像検索システム。

【請求項2】 前記属性リスト生成部は、属性リストのデータ構造を定義したシンタックスに基づいて当該属性リストを生成し、前記画像検索部は、前記属性リストのシンタックスに従って属性リストを解析し、検索することを特徴とする請求項1記載の画像検索システム。

【請求項3】 前記特徴記述子生成部は、特徴記述子のデータ構造を定義したシンタックスに基づいて当該特徴記述子を生成し、前記画像検索部は、前記特徴記述子のシンタックスに従って特徴記述子を解析し、検索することを特徴とする請求項1または請求項2記載の画像検索システム。

【請求項4】 前記特徴記述子生成部は、フレーム単位で特徴量を抽出し、複数のフレームをまとめたビデオセグメント単位で特徴記述子を生成することを特徴とする請求項1～3のいずれかに記載の画像検索システム。

【請求項5】 前記特徴記述子生成部は、イントラフレームとインターフレームとからなる圧縮映像データを入力画像データとし、

イントラフレームの場合、所定の符号化単位領域における各画素値の平均値マトリクスを抽出してビデオセグメント中に含まれるイントラフレーム分だけ累積し、その累積値をイントラフレーム数により平均した値と標準偏差とをイントラフレームの特徴記述子として生成する一方、

インターフレームの場合、所定の符号化単位領域における各画素の動きベクトルマトリクスを抽出してフレーム内の動きベクトルの平均値を求め、その平均値を閾値処理してゼロランの長さを分類して抽出し、前記平均値と前記ゼロランの長さの分類とをそれぞれビデオセグメント中に含まれるインターフレーム数により平均した値をインターフレームの特徴記述子として生成する、ことを特徴とする請求項4記載の画像検索システム。

【請求項6】 前記特徴記述子生成部は、非圧縮映像データを入力画像データとし、前記非圧縮映像データデータから特徴量を抽出し特徴記述子を生成し、前記画像情報蓄積部は、前記非圧縮映像データが所定の

2

映像圧縮方式に基づいて圧縮された後の圧縮映像データと、前記特徴記述子セットとを対応付けて蓄積することとを特徴とする請求項1～4のいずれかに記載の画像検索システム。

【請求項7】 前記画像検索部は、前記画像情報蓄積部を検索し、指定された特徴記述子に適合する画像データを出力する際、予め記憶していた所定の条件式を読み出し、その所定の条件式に基づいて当該指定された特徴記述子に対し適合しているか否かを判定することを特徴とする請求項1～5のいずれかに記載の画像検索システム。

【請求項8】 入力画像データは、監視カメラにより記録された監視画像であることを特徴とする請求項1～7のいずれかに記載の画像検索システム。

【請求項9】 入力画像データは、蓄積されたビデオメールの画像であることを特徴とする請求項1～7のいずれかに記載の画像検索システム。

【請求項10】 入力画像データは、蓄積された放送番組の画像であることを特徴とする請求項1～7のいずれかに記載の画像検索システム。

【請求項11】 入力画像データは、ビデオカメラにより記録された映像であることを特徴とする請求項1～7のいずれかに記載の画像検索システム。

【請求項12】 入力画像データから画像特徴量を抽出し特徴記述子を生成し、生成された特徴記述子を入力画像データと対応づけて蓄積すると共に、入力画像データに付随して入力する属性情報に基づいて属性リストを生成し、

属性情報に関する検索条件が入力すると前記属性リストを検索して当該検索条件に適合する属性情報を出力すると共に、当該出力された属性情報を参照して特徴記述子に関する検索条件が入力すると前記蓄積された画像データを検索して当該検索条件に適合する画像データを出力する、ことを特徴とする画像検索方法。

【請求項13】 入力画像データから画像特徴量を抽出し特徴記述子を生成する特徴記述子生成部と、入力画像データの時空間構造に関連付けて前記特徴記述子を検索用記述データに記述する検索用記述データ生成部と、

検索用記述データを対応する入力画像データとともに蓄積する画像情報蓄積部と、

第1の検索処理部と、

第2の検索処理部と、

ユーザインターフェース部からなり、

前記第2の検索処理部は、前記ユーザインターフェース部を介したユーザからの検索要求を検索要求情報として前記第1の検索処理部へ送信すると共に、前記第1の検索処理部から送信される検索結果を受信して前記ユーザインターフェース部を介しユーザに提示し、

(3)

3  
前記第1の検索処理部は、前記第2の検索処理部から送信される前記検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子を取り出し、取り出した前記特徴記述子に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、  
ことを特徴とする画像検索システム。

【請求項14】 さらに、前記特徴記述子生成部によって生成された特徴記述子の信頼度を算出する特徴記述子信頼度算出部を備え、

前記検索用記述データ生成部は、入力画像データの時空間構造に関連付けて前記特徴記述子および信頼度を検索用記述データに記述し、

前記第1の検索処理部は、前記第2の検索処理部から送信される前記検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子および信頼度を取り出し、取り出した前記特徴記述子および信頼度に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、

ことを特徴とする請求項13記載の画像検索システム。

【請求項15】 前記第1の検索処理部は、適合判定処理の際、特徴記述子の信頼度に基づいて、特徴記述子による適合判定の必要性を評価し、適合性判定を行う必要がないと判断した場合、適合判定処理をスキップすることを特徴とする請求項14記載の画像検索システム。

【請求項16】 さらに、前記特徴記述子信頼度算出部によって算出された信頼度に基づいて各特徴記述子の検索使用順序を決定する特徴記述子検索順序決定部を備え、

前記検索用記述データ生成部は、入力画像データの時空間構造に関連付けて前記特徴記述子、信頼度および検索使用順序を検索用記述データに記述し、

前記第1の検索処理部は、前記第2の検索処理部から送信される前記検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子、信頼度および検索使用順序を取り出し、取り出した前記特徴記述子、信頼度および検索使用順序に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、

ことを特徴とする請求項14または請求項15記載の画像検索システム。

【請求項17】 前記第2の検索処理部は、前記ユーザインターフェース部を介したユーザからの検索要求として各特徴記述子の検索使用順序を含む検索要求情報を前

4  
記第1の検索処理部へ送信し、

前記第1の検索処理部は、前記第2の検索処理部から送信される検索要求情報中の検索使用順序に従って、特徴記述子の検索順序を更新して、ユーザからの検索要求に見合う検索処理を行うことを特徴とする請求項13～請求項16のいずれかに記載の画像検索システム。

【請求項18】 前記第1の検索処理部は、適合判定処理を行う際、各特徴記述子の信頼度に基づいて、検索使用順序で定められる各検索ステップにおける検索結果候補数を決定して適合判定処理を行うことを特徴とする請求項16または請求項17記載の画像検索システム。

【請求項19】 さらに、

ビデオデータ再生サーバと、

ビデオデータ復号再生部とを備え、

前記入力画像データは、ビデオデータであり、

前記第1の検索処理部は、検索によって特定された各ビデオデータを代表する各キー画像データを検索結果として前記第2の検索処理部へ送信し、

前記第2の検索処理部は、前記第1の検索処理部からの検索結果としての前記各キー画像データを受信してユーザインターフェース部を介しユーザに提示させ、

前記ビデオ再生サーバは、ユーザインターフェース部から前記各キー画像のうちユーザによって特定のキー画像を選択した再生要求を受けると、画像情報蓄積部からその選択されたキー画像を代表とするビデオデータを読み出して前記ビデオデータ復号再生部へ送信し、

前記ビデオデータ復号再生部は、前記ビデオ再生サーバから送信されたビデオデータを受信して復号し再生する、

ことを特徴とする請求項13～請求項18のいずれかに記載の画像検索システム。

【請求項20】 入力画像データから画像特徴量を抽出し特徴記述子を生成すると共に、入力画像データの時空間構造に関連付けて前記特徴記述子を検索用記述データに記述して、検索用記述データを対応する入力画像データとともに蓄積し、

ユーザからの検索要求に従って前記蓄積された入力画像データの検索用記述データを解析して前記特徴記述子を取り出し、取り出した前記特徴記述子に基づいて適合判定処理を行って検索結果を求め、その検索結果をユーザに提示する、

ことを特徴とする画像検索方法。

【請求項21】 さらに、入力画像データは、ビデオデータであり、

検索によって特定された各ビデオデータを代表する各キー画像データを検索結果としてユーザに提示し、

前記各キー画像のうちユーザによって特定のキー画像を選択した再生要求を受けると、その選択されたキー画像を代表とするビデオデータを読み出して復号し再生する、

50

(4)

5

ことを特徴とする請求項20記載の画像検索方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、アナログまたはデジタルで記録された映像、静止画などの画像データから特徴量や属性情報を抽出し、抽出された特徴量や属性情報を利用して画像データの検索を行う画像検索システムおよび画像検索方法に関する。

【0002】

【従来の技術】図14に、従来の画像検索処理システムの一例として、電子情報通信学会論文誌D-II 1996年4月号 (Vol. 79-D-II No. 4, pp. 476-483) に記載されたシステムの構成を示す。ここでは、静止画像を対象とし、画像を前処理部101の領域分割部103により各セグメントに領域分割し、分割された各セグメントに対してキーワードが付与される。キーワードには概念キーワードとシーン記述キーワードの2種類が用意されている。概念キーワード抽出部104は、色情報に対してあらかじめ割り当てられている概念キーワードを用いて、各セグメントの色および特徴量から概念キーワード108を得るものである。シーン記述キーワード記述部105は、各セグメントの画像特徴量から「位置」「色」「形状」「大きさ」「向き」などに関する述語を得るものであり、ユーザ107によってあらかじめ定義されている述語から選択して記述する操作106を必要とし、シーン記述キーワード109を出力する。検索ツール102は、あらかじめ概念キーワードやシーン記述キーワードを用意しておき、ユーザ111は用意されているキーワードから選択を行う。特徴同定部110は、ユーザが選択したキーワード112と、各画像に付与されたキーワード108、109から特徴量の同定を行い、画像検索処理を実行する。

【0003】

【発明が解決しようとする課題】しかし、前述の画像検索処理システムでは、ユーザが選択した概念キーワードやシーン記述キーワード等のキーワードと、各画像に付与されたキーワードから特徴量の同定を行い、その特徴量に基づいて画像検索処理を実行するようにしていたため、すべての画像を特徴量のみにより検索することになり、検索に時間がかかっていた。

【0004】また、前述の画像検索処理システムでは、各キーワードの記述方法や、記憶方法等について考慮されていないため、画像サーバと、クライアントである検索ツール等とは、1対1に対応している必要があり、ネットワークを介して多くのユーザが様々な検索ツールを用いて画像の検索を行うシステムを構築することができなかった。

【0005】また、静止画像のみを対象としているため、動画像を検索することが困難であった。

【0006】そこで、本発明は、このような問題に鑑みてなされたもので、効率良く画像検索処理を実行できる

6

画像検索システムおよび画像検索方法を提供することを第1の目的とする。

【0007】また、本発明は、検索のキーワードを共通のシンタックスで記述して生成することにより、ネットワーク上に分散する複数の画像サーバに依存しない画像検索システムおよび画像検索方法を提供することを第2の目的とする。

【0008】また、本発明は、動画像からキーワードを抽出するときに、フレーム単位ではなく、複数のフレームをまとめたビデオセグメント単位で特徴量を抽出することにより、動画像の検索を容易に行うことのできる画像検索システムおよび画像検索方法を提供することを第3の目的とする。

【0009】

【課題を解決するための手段】前記課題を解決するため、本発明では、入力画像データから画像特徴量を抽出し特徴記述子を生成する特徴記述子生成部と、生成された特徴記述子を入力画像データと対応づけて蓄積する画像情報蓄積部と、入力画像データに付随して入力する属性情報に基づいて属性リストを生成する属性リスト生成部と、属性情報に関する検索条件が入力すると前記属性リストを検索して当該検索条件に適合する属性情報を入力すると共に、特徴記述子に関する検索条件が入力すると前記画像情報蓄積部を検索して当該検索条件に適合する画像データを出力する画像検索部と、を備えることを特徴とする。

【0010】特に、前記属性リスト生成部は、属性リストのデータ構造を定義したシンタックスに基づいて当該属性リストを生成し、前記画像検索部は、前記属性リストのシンタックスに従って検索することを特徴とする。

【0011】また、前記特徴記述子生成部は、特徴記述子のデータ構造を定義したシンタックスに基づいて当該特徴記述子を生成し、前記画像検索部は、前記特徴記述子のシンタックスに従って検索することを特徴とする。

【0012】また、前記特徴記述子生成部は、フレーム単位で特徴量を抽出し、複数のフレームをまとめたビデオセグメント単位で特徴記述子を生成することを特徴とする。

【0013】また、前記特徴記述子生成部は、イントラフレームとインターフレームとからなる圧縮映像データを入力画像データとし、イントラフレームの場合、所定の符号化単位領域における各画素値の平均値マトリクスを抽出してビデオセグメント中に含まれるイントラフレーム分だけ累積し、その累積値をイントラフレーム数により平均した値と標準偏差とをイントラフレームの特徴記述子として生成する一方、インターフレームの場合、所定の符号化単位領域における各画素の動きベクトルマトリクスを抽出してフレーム内の動きベクトルの平均値を求め、その平均値を閾値処理してゼロランの長さを分類して抽出し、前記平均値と前記ゼロランの長さの分類

(5)

7

とをそれぞれビデオセグメント中に含まれるインターフレーム数により平均した値をインターフレームの特徴記述子として生成する、ことを特徴とする。

【0014】また、前記特徴記述子生成部は、非圧縮映像データを入力画像データとし、前記非圧縮映像データから特徴量を抽出し特徴記述子を生成し、前記画像情報蓄積部は、前記非圧縮映像データが所定の映像圧縮方式に基づいて圧縮された後の圧縮映像データと、前記特徴記述子セットとを対応付けて蓄積することを特徴とする。

【0015】また、前記画像検索部は、前記画像情報蓄積部を検索し、指定された特徴記述子に適合する画像データを出力する際、予め記憶していた所定の条件式を読み出し、その所定の条件式に基づいて当該指定された特徴記述子に対し適合しているか否かを判定することを特徴とする。

【0016】また、入力画像データは、監視カメラにより記録された監視画像であることを特徴とする。

【0017】また、入力画像データは、蓄積されたビデオメールの画像であることを特徴とする。

【0018】また、入力画像データは、蓄積された放送番組の画像であることを特徴とする。

【0019】また、入力画像データは、ビデオカメラにより記録された映像であることを特徴とする。

【0020】また、次の発明では、入力画像データから画像特徴量を抽出し特徴記述子を生成する特徴記述子生成部と、入力画像データの時空間構造に関連付けて前記特徴記述子を検索用記述データに記述する検索用記述データ生成部と、検索用記述データを対応する入力画像データとともに蓄積する画像情報蓄積部と、第1の検索処理部と、第2の検索処理部と、ユーザインターフェース部からなり、前記第2の検索処理部は、前記ユーザインターフェース部を介したユーザからの検索要求を検索要求情報として前記第1の検索処理部へ送信すると共に、前記第1の検索処理部から送信される検索結果を受信して前記ユーザインターフェース部を介しユーザに提示し、前記第1の検索処理部は、前記第2の検索処理部から送信される前記検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子を取り出し、取り出した前記特徴記述子に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、ことを特徴とする。

【0021】さらに、前記特徴記述子生成部によって生成された特徴記述子の信頼度を算出する特徴記述子信頼度算出部を備え、前記検索用記述データ生成部は、入力画像データの時空間構造に関連付けて前記特徴記述子および信頼度を検索用記述データに記述し、前記第1の検索処理部は、前記第2の検索処理部から送信される前記

8

検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子および信頼度を取り出し、取り出した前記特徴記述子および信頼度に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、ことを特徴とする。

【0022】また、前記第1の検索処理部は、適合判定処理の際、特徴記述子の信頼度に基づいて、特徴記述子による適合判定の必要性を評価し、適合性判定を行う必要がないと判断した場合、適合判定処理をスキップすることを特徴とする。

【0023】さらに、前記特徴記述子信頼度算出部によって算出された信頼度に基づいて各特徴記述子の検索使用順序を決定する特徴記述子検索順序決定部を備え、前記検索用記述データ生成部は、入力画像データの時空間構造に関連付けて前記特徴記述子、信頼度および検索使用順序を検索用記述データに記述し、前記第1の検索処理部は、前記第2の検索処理部から送信される前記検索要求情報に従って前記画像情報蓄積部に蓄積された入力画像データの前記検索用記述データを解析して前記特徴記述子、信頼度および検索使用順序を取り出し、取り出した前記特徴記述子、信頼度および検索使用順序に基づいて適合判定処理を行って検索結果を求め、その検索結果を前記第2の検索処理部へ送信し、前記ユーザインターフェース部を介しユーザに提示させる、ことを特徴とする。

【0024】また、前記第2の検索処理部は、前記ユーザインターフェース部を介したユーザからの検索要求として各特徴記述子の検索使用順序を含む検索要求情報を前記第1の検索処理部へ送信し、前記第1の検索処理部は、前記第2の検索処理部から送信される検索要求情報中の検索使用順序に従って、特徴記述子の検索順序を更新して、ユーザからの検索要求に見合う検索処理を行うことを特徴とする。

【0025】また、前記第1の検索処理部は、適合判定処理を行う際、各特徴記述子の信頼度に基づいて、検索使用順序で定められる各検索ステップにおける検索結果候補数を決定して適合判定処理を行うことを特徴とする。

【0026】さらに、ビデオデータ再生サーバと、ビデオデータ復号再生部とを備え、前記入力画像データは、ビデオデータであり、前記第1の検索処理部は、検索によって特定された各ビデオデータを代表する各キー画像データを検索結果として前記第2の検索処理部へ送信し、前記第2の検索処理部は、前記第1の検索処理部からの検索結果としての前記各キー画像データを受信してユーザインターフェース部を介しユーザに提示させ、前記ビデオ再生サーバは、ユーザインターフェース部から前記各キー画像のうちユーザによって特定のキー画像を

9

選択した再生要求を受けると、画像情報蓄積部からその選択されたキー画像を代表とするビデオデータを読み出して前記ビデオデータ復号再生部へ送信し、前記ビデオデータ復号再生部は、前記ビデオ再生サーバから送信されたビデオデータを受信して復号し再生する、ことを特徴とする。

【0027】また、次の発明では、入力画像データから画像特徴量を抽出し特徴記述子を生成すると共に、入力画像データの時空間構造に関連付けて前記特徴記述子を検索用記述データに記述して、検索用記述データに対応する入力画像データとともに蓄積し、ユーザからの検索要求に従って前記蓄積された入力画像データの検索用記述データを解析して前記特徴記述子を取り出し、取り出した前記特徴記述子に基づいて適合判定処理を行って検索結果を求め、その検索結果をユーザに提示する画像検索方法であることを特徴とする。

【0028】さらに、入力画像データは、ビデオデータであり、検索によって特定された各ビデオデータを代表する各キー画像データを検索結果としてユーザに提示し、前記各キー画像のうちユーザによって特定のキー画像を選択した再生要求を受けると、その選択されたキー画像を代表とするビデオデータを読み出して復号し再生する、ことを特徴とする。

【0029】

【発明の実施の形態】実施の形態1. 本実施の形態1では、ネットワーク上に散在する複数、すなわち1つ以上のデータベースに蓄積された画像データからユーザが所望の画像データを検索して利用することができる画像検索システムについて説明する。画像データは利用方法によって、圧縮静止画像（JPEG画像等）、圧縮動画（MP  
EG画像等）、非圧縮画像いづれでもよく、画像フォーマットについては特に限定しない。本実施の形態1では、本発明の一例として、インターネット上のWWW（World Wide Web）により、画像情報の発信や収集を行うことを想定するが、本発明はこれに限るものではなく、有線または無線通信によるサービスや、放送網によるサービス等に本システムを適用することもできる。

【0030】図1は、本実施の形態1における画像検索システムの構成を説明するブロック図である。以下、「クラス」をデータの定義、「インスタンス」をクラス定義に基づき値として具象化したデータそのもの、例えばファイル、変数などの意味で用いる。

【0031】図1において、1は画像データベース生成部、2は画像に関連する属性を示すテキスト情報の属性情報および画像データ、3は画像データから所定の特徴量セットを抽出する特徴抽出部、4は他の画像データベース生成部1あるいはサーバと共通の特徴量記述データクラスによる定義に基づき特徴量セットからインスタンスセットである特徴記述子セットを生成する特徴記述子  
セット生成部、5は画像データと特徴記述子セットとを

(6)

10

対で蓄積する画像情報蓄積部である。

【0032】また、図1において、6は他の画像データベース生成部1あるいはサーバと共通のテキスト情報記述データクラスによる定義に基づき属性情報からインスタンスセットである属性リストを生成する属性リスト生成部、7は属性リストを蓄積する属性リスト蓄積部である。さらに8は画像検索部であり、9は検索対象の画像データのカテゴリやキー画像から検索したい画像を選択するユーザインターフェース部、10はユーザインターフェース部で指示された検索条件に基づき検索処理を行う検索処理部、11は検索されたキー画像および検索画像を表示する表示部である。なお、本実施の形態1において、属性リストや特徴記述子を総称してメタデータと定義する。

【0033】また、図1では、特徴抽出部3、特徴記述子セット生成部4、画像情報蓄積部5、属性リスト生成部6および属性リスト蓄積部7からなる画像データベース生成部1と、画像検索部8の検索処理部10がサーバを構成し、画像検索部8のユーザインターフェース部9および表示部11がクライアントを構成して、サーバとクライアントとにより本発明の画像検索システムを構築しているが、本発明では、このような構成に限らず、クライアント側に検索処理部10を持ってきたり、あるいはサーバ自身にもユーザインターフェース部9および表示部11のクライアントの機能を付加する等するようにしても勿論よい。また、画像データベース生成部1ないしはサーバ、および画像検索部8ないしはクライアントは、それぞれ、複数存在し、かつ、インターネット等のネットワークを介して接続されており、その複数の画像データベース生成部1ないしはサーバでは、上述の通り、共通の特徴量記述データクラスおよびテキスト情報記述データクラスに基づき特徴記述子セットおよび属性リストを生成する。

【0034】次に、各部の処理を詳述する。

#### （1）画像データベース生成部1の処理

図2に、画像データベース生成部1の処理をフローチャートにより示す。ここで、新規の画像データを画像データベースに登録する場合を考える。画像データベースの登録は、主に2つの処理からなる。1つは画像データに附随する画像属性を示すテキスト情報から属性リストを生成し、属性リスト蓄積部7へ蓄積する処理である。もう1つは画像データから特徴量を抽出して、特徴記述子を生成し、画像データと特徴記述子とをデータベースである画像情報蓄積部5へ蓄積する処理である。次にそれぞれの処理の詳細について説明する。

【0035】先に、属性リストを生成し登録するまでの処理（ステップS1～S3）について詳述する。まず、画像情報に附随して入力される画像属性を示すテキスト情報を抽出する（ステップS1）。画像属性を示すテキスト情報としては、例えば、画像データがビデオメール



(7)

11

の動画データである場合には、動画データと一緒に送信されるビデオメールの送信者、送信日時、タイトル等の情報である。これらの情報は、ビデオメールの動画データの作成者や作成日などを示しており、動画データの属性を示すテキスト情報である。また、別の例では、画像データが、放送局から送信されるテレビ放送等の映像プログラムである場合には、画像属性を示すテキスト情報としては、映像プログラムと一緒に送信される番組情報（出演者、演出者、タイトル、放送日など）が考えられる。ホームビデオやデジタル写真などを家庭内サーバに蓄積する例を想定する場合は、撮影日や被写体に関する情報をユーザがテキスト情報として直接入力したり、ビデオやカメラなどの装置からサーバにテキストの形式で転送されることも考えられる。

【0036】抽出されたテキスト情報は、属性リスト生成部6において属性リストの形式にインスタンス化される（ステップS2）。属性リストのデータ形式としては、例えばXML（eXtensible Markup Language）により記述された文書がある。

【0037】図3に、画像属性のテキスト情報からテキスト情報のインスタンスであるXML文書を生成する処理の一例を示す。図3の例では、属性情報としてビデオメールに付属するテキスト情報を抽出して、抽出したテキスト情報をXML言語で記述した文書を生成している。このXML文書は、属性リストの構成単位となるデータであり、個々のビデオメールに対して作成されたこれらの文書をひとつのリストデータの形態に統合したものが属性リストである。XML文書の記述には、文書構造を規定した定義ファイル（DTD（Document Type Definition））が必要である。DTDは、文書構造に含まれる要素と要素間の関係（どの要素が別の要素の中に出現するか等）、要素の出現順序等の規定を定めたものである。

【0038】図3に示すXML文書生成処理では、DTDは、TextInfoという要素に、Category、Author、Date、Time、Title、Locatorの各要素が含まれていて、これらの各要素が文字型のデータであることを示している。抽出されたテキスト情報はこのDTDに従ってXML文書に変換される。図3のXML文書の例における各要素の値は、要素の開始を表すタグ<aaa>と、要素の終わりを示すタグ</aaa>で囲まれており、aaaはDTDで定義された要素名を示している。すなわち、<TextInfo>〜</TextInfo>で囲まれる単位はひとつのビデオメールに関する属性を記述しており、検索対象となるすべてのビデオメールについてこれらのデータが統合された属性リストを持っておけば、<TextInfo>〜</TextInfo>の単位で属性による個々のビデオメールの検索を行うことが可能になる。

【0039】尚、属性リストのデータ形式は、これに限ったものではなく、予め定義されたデータ構造に基づいて記述されたものであればよい。例えば、属性リストのデータ形式として、XML文書として記述する以外に、ビ

12

ットストリーム型のバイナリデータでもよい。この場合、XML文書と同様に、バイナリデータ構造を定義するシンタックス、すなわち各要素の並び方と各要素に割り当てられたビット数等を規定したものが必要となる。

【0040】以上のようにして生成された属性リストは、属性リスト蓄積部7に蓄積される（ステップS3）。なお、画像データと、その属性情報であるテキスト情報とが異なるデータベースに蓄積される場合、属性リストは、画像データが蓄積されたアドレスを含んで属性リスト蓄積部7に蓄積される。たとえば、WWWにおいては、画像データが位置するネットワークアドレス（URL（Universal Resource Locator））を指定すればよい。図3の例においては、属性リストであるXML文書に、“Locator”という画像データのURLを指定するための要素が含まれており、各々のビデオメールの属性情報に対して、対応する画像データのアドレスを指定することができる。

【0041】次に、画像の特徴量を抽出して特徴量記述子を生成し登録するまでの処理（ステップS4〜S6）について詳述する。本実施の形態1における特徴記述子生成処理では、入力して蓄積される画像データは、MPEGやH.261/H.263等の所定の符号化方式によって圧縮符号化されたビデオ（動画）ビットストリームであるものとする。

【0042】まず、特徴抽出部3において、画像ビットストリームより画像の特徴量を抽出する（ステップS4）。本実施の形態1においては、特徴量として、圧縮ビデオビットストリーム中の動き情報、イントラ符号化フレームの輝度、色差それぞれの直流成分のデータを抽出する。圧縮ビデオビットストリームは、MPEGやH.261/H.263などの国際標準符号化方式によるものを想定しており、これらの方式においてはビデオフレーム信号は、図4に示すようにマクロブロックという単位で動き補償予測（以下、インターという）／フレーム内（以下、イントラという）適応符号化されている。マクロブロックは、輝度信号16x16画素、色差信号8x8画素x2の画像データから構成される。特に、1フレームのすべてのマクロブロックをイントラ符号化する場合をイントラフレームと呼び、この場合、当該フレームの画像データはそれ自身の圧縮データのみで伸長・復元可能である。このようなフレームにおいては、イントラ符号化マクロブロックの各信号の直流成分が画像全体の概略を表現するデータであると考えられる（図4中、左側最下段に図示）。イントラフレームは、図4に示すようにランダムアクセスなどの要求条件から通常ビデオ信号中に周期的に挿入されることが多い。

【0043】一方、動き補償予測によって予測残差信号を符号化するインター符号化マクロブロックでは、動き予測によって近似しきれなかったデータのみを符号化するため、その符号化情報中、最も本質的なデータはマク

(8)

13

ロブロックの大方の画像情報を近似する動きベクトルであると言える(図4中、右側最下段に図示)。以上の観点から、特徴抽出部3では、画像ビットストリームより画像の特徴量として、イントラフレームにおける各マクロブロックの輝度・色差信号の直流成分と、インター符号化マクロブロックの動きベクトルとを抽出するものとする。

【0044】以下の説明のため、所定の期間のフレーム群をビデオセグメントと呼ぶこととする。例えば、最も単純なビデオセグメントはイントラフレームから始まり、次のイントラフレームの直前のフレームで終了する単位と考えられる。または、より長い期間でイントラフレームからイントラフレーム直前フレームまでのフレーム群を一まとめにしたフレームセットと考えることもできる。ビデオセグメントの長さはビデオセグメント内のコンテンツの意味内容に従い、任意に設定されてよい。

【0045】本実施の形態1では、前記特徴抽出部3によって抽出された特徴量セットに基づいて、特徴記述子セット生成部4において、ビデオセグメントを記述するための特徴記述子セットを生成する(ステップS5)。以下、特徴量セットの抽出手順と特徴記述子セットの生成手順について詳しく説明する。

【0046】図5は、図2におけるS4の特徴抽出処理およびS5の特徴記述子生成処理を含めたトータルの特徴記述子セット生成手順を詳細に示すフローチャートである。尚、特徴抽出部3に入力する画像データであるビデオビットストリームは、あらかじめ切り出されたビデオセグメントを単位とするビットストリームであってもよいし、複数のビデオセグメントに分けられる前のビデオプログラムのような長時間に渡るビットストリームでもよい。後者の場合、特徴抽出部3において、ビデオセグメントの単位を外部からの指示、もしくは適当な内部での識別処理により特定し、その単位で特徴抽出・特徴記述子セットを抽出するよう処理すればよい。以下では、基本となるビデオセグメント単位のビットストリームが入力される場合の処理について述べる。ビデオセグメントは、上述の通り、イントラフレームで始まる一連のフレームセットから構成されているものとする。特徴抽出は、フレームを単位として、イントラフレームとインターフレームとに分けて実施する。イントラフレームの場合は輝度・色差信号の直流成分、インターフレームの場合は動きベクトルを抽出する。

【0047】具体的には、まず、ステップS7においてフレームのヘッダ情報をサーチして当該フレームがイントラフレーム(I-frame)の場合は(ステップS7 “Y”)、ステップS8によりマクロブロックごとに各画素値の平均値マトリクスである輝度・色差信号の直流成分のデータを抽出する。図5において、マクロブロックの位置はフレーム上のラスタスキャン順でnで表し、位置nのマクロブロックの輝度信号の直流成分をDC

14

Y[n]、色差信号の直流成分をDC<sub>U</sub>[n]、DC<sub>V</sub>[n]と表す。nの最大値はフレームの水平・垂直画素数で与えられる。MPEG, H. 26xシリーズなどの国際標準ビデオ符号化方式では、直流成分はマクロブロックを構成する各8x8画素のサブブロック単位で実施されるDCT(離散コサイン変換)の(0,0)位置のDC成分で表されるため、マクロブロック単位にこのデータをサーチして取り出せばよい。ただし、輝度信号については1マクロブロックあたり4つのサブブロックがあるため、輝度信号の直流成分としては4つのサブブロックのDC成分を平均した値とする。

【0048】次いでステップS9において、イントラフレーム対応の特徴量を累積する。このステップは、特徴記述子セット生成部4にて行われる。具体的には、DC Y[n]、DC<sub>U</sub>[n]、DC<sub>V</sub>[n]をそれぞれ下式1に基づいて累積する。

【0049】

【数1】

$$\begin{aligned} \text{avg}Y_{\text{acc}}[n] &= \text{avg}Y_{\text{pre}}[n] + \text{DC}_Y[n]; & \text{avg}_{\text{pre}}Y[n] &= \text{avg}_{\text{acc}}Y[n] \\ \text{avg}U_{\text{acc}}[n] &= \text{avg}U_{\text{pre}}[n] + \text{DC}_U[n]; & \text{avg}_{\text{pre}}U[n] &= \text{avg}_{\text{acc}}U[n] \\ \text{avg}V_{\text{acc}}[n] &= \text{avg}V_{\text{pre}}[n] + \text{DC}_V[n]; & \text{avg}_{\text{pre}}V[n] &= \text{avg}_{\text{acc}}V[n] \\ \text{std}Y_{\text{acc}}[n] &= \text{std}Y_{\text{pre}}[n] + \text{DC}_Y[n] * \text{DC}_Y[n]; & \text{std}Y_{\text{pre}}[n] &= \text{std}Y_{\text{acc}}[n] \\ \text{std}U_{\text{acc}}[n] &= \text{std}U_{\text{pre}}[n] + \text{DC}_U[n] * \text{DC}_U[n]; & \text{std}U_{\text{pre}}[n] &= \text{std}U_{\text{acc}}[n] \\ \text{std}V_{\text{acc}}[n] &= \text{std}V_{\text{pre}}[n] + \text{DC}_V[n] * \text{DC}_V[n]; & \text{std}V_{\text{pre}}[n] &= \text{std}V_{\text{acc}}[n] \end{aligned}$$

【0050】尚、avgY<sub>pre</sub>[n]、avgU<sub>pre</sub>[n]、avgV<sub>pre</sub>[n]、stdY<sub>pre</sub>[n]、stdU<sub>pre</sub>[n]、stdV[n]の各値は、抽出処理開始時にゼロにリセットしておく。

【0051】イントラフレーム中のすべてのnについて抽出および累積の処理が終了したら(ステップS10 “Y”)、ビデオセグメント中のイントラフレーム出現数カウンタN<sub>I</sub>をインクリメントし、次のフレームの処理に移る(ステップS11 “N”)。ビデオセグメントの終了であれば(ステップS11 “Y”)、特徴記述子セット生成処理のステップS17に移る。

【0052】次にインターフレームの特徴量抽出について述べる。ステップS7でイントラフレームでないと判断されると(ステップS7 “N”)、ステップS12で前(片)方向予測のインターフレーム(P-frame)かどうかを調べる。ここで、P-frameとしたのは、MPEGでは、インターフレームとして、P-frameの他に、時間的に過去および未来のフレームの両方を参照して動き補償予測を行う両方向予測フレーム(B-frame)や、スプライトと呼ばれる全景画像からの変形合成で復号画像を復元するスプライト予測フレーム(ステップS-frame, S-VOP)などのフレーム符号化タイプがあり、これらとの区別を行うためである。ここでは、インターフレームとして、過去のフレームのみを参照して前(片)方向の動き補償予測を行うインターフレーム(P-frame)のみを対象とする。したがって、ステップS12でP-frameでないと判断された場合は(ステップS12 “N”)、特徴量の抽出を行わず、次のフレームの処理に移行する。



(9)

15

【0053】ところで、ステップS12でインターフレーム (P-frame) であると判断された場合には (ステップS12 “Y”)、特徴抽出部3において、フレーム内のすべてのマクロブロックの動きベクトルMV[n]を抽出する (ステップS13)。ここで、MV[n]は下式2で表される値とする。

【0054】

【数2】

$$MV[n] = \sqrt{x_n^2 + y_n^2}$$

$x_n$ : 位置nにおける動きベクトルの水平成分  
 $y_n$ : 位置nにおける動きベクトルの垂直成分

【0055】そして、ステップS14ですべてのマクロブロックについてMV[n]の抽出が終了したと判断されたら (ステップS14 “Y”)、ステップS15において、当該フレームに対応するインターフレームの特徴記述子セットを生成する。この処理は、特徴記述子セット生成部4にて行われる。

【0056】図6に、図5におけるステップS15の特徴記述子セット処理の具体的な生成手順をフローチャートにより示す。まず、ステップS13の処理により1フレ \* 20

$N_s$ : フレーム水平画素数の1/3以下の長さをもつゼロラン(short run)の数  
 $N_m$ : フレーム水平画素数の1/3以上2/3の長さをもつゼロラン(middle run)の数  
 $N_l$ : フレーム水平画素数の2/3以上の長さをもつゼロラン(long run)の数

【0062】尚、図7からわかるように、short runが多いフレームでは動きの複雑度が大きく、広い範囲に動きが分布していることを示し、long runが多いフレームでは動きが局在化されていることを示す。

【0063】インターフレーム中のすべてのnについて抽出および累積の処理が終了し (ステップS14 “Y”)、ビデオセグメント中のインターフレーム出現数カウンタNpをインクリメントし、次のフレームの処理に移る (ステップS16 “N”)。ビデオセグメントの終了であれば (ステップS16 “Y”)、特徴記述子セット生成処理のステップS17に移る。

【0064】以上の結果、ビデオセグメント中の各インターフレームについてその特徴記述子セット ( $C_{Np}$ ,  $N_{sr}$ ,  $N_{mr}$ ,  $N_{lr}$ ) が得られる。これらはステップS17のビデオセグメントの特徴記述子セット生成のために、インターフレームのたびに累積される。

【0065】以上のように、各フレームごとの特徴抽出もしくは特徴記述子の生成をビデオセグメント中のすべてのフレームについて実施した後、最後にステップS17において、ビデオセグメントの特徴記述子セットを生成する。この処理も特徴記述子セット生成部4にて行う。

【0066】具体的には、イントラフレームについて、累積された特徴量  $avgY_{sum}[n]$ ,  $avgU_{sum}[n]$ ,  $avgV_{sum}[n]$ ,  $stdY_{sum}[n]$ ,  $stdU_{sum}[n]$ ,  $stdV_{sum}[n]$  と、ビデオセグメント内のイントラフレーム数  $N_I$  を用いて、下式5の特徴記述子セット ( $avgY[n]$ ,  $avgU[n]$ ,  $avgV[n]$ ,  $stdY[n]$ ,  $stdU[n]$ ,  $stdV[n]$ ) を生成する。

16

\* 一ム分蓄積されたMV[n]について、フレーム内の平均動き量として下式3の値を求める (ステップS15-1)。

【0057】

【数3】

$$C_{Np} = \frac{1}{N} \sum_{n=0}^{N-1} MV[n]$$

【0058】ただし、Nはフレーム内マクロブロック総数を示す。つぎにステップS15-2において、各MV[n]に対して下式4の閾値処理を施す。

10 【0059】

【数4】

$$MV'[n] = \begin{cases} MV[n]; & \text{if } MV[n] \geq C_{Np} \\ 0; & \text{if } MV[n] < C_{Np} \end{cases}$$

【0060】こうして得られた新しい動きベクトル分布に対し、MV'[n]をラスタスキャン順にスキャンし、以下の動き分布を特徴づける量を求める (ステップS15-3)。

【0061】

【0067】

【数5】

$$\begin{aligned} avgY[n] &= avgY_{sum}[n] / N_I \\ avgU[n] &= avgU_{sum}[n] / N_I \\ avgV[n] &= avgV_{sum}[n] / N_I \\ stdY[n] &= \sqrt{\frac{stdY_{sum}[n]}{N_I} - (avgY[n])^2} \\ stdU[n] &= \sqrt{\frac{stdU_{sum}[n]}{N_I} - (avgU[n])^2} \\ stdV[n] &= \sqrt{\frac{stdV_{sum}[n]}{N_I} - (avgV[n])^2} \end{aligned}$$

【0068】以上の結果、イントラフレームに関して、 $n \times 6$ 個の特徴記述子セットが生成される。これらの特徴記述子セットの意味を図8を用いて説明する。

【0069】まず、各フレームにおいて各マクロブロックの直流成分を求めることは、図8における各マクロブロック位置の画像データの平均的な明るさと色とを時間方向にプロットした波形を得ることと同値である。すなわち、 $avgY[n]$ の時間系列は、各マクロブロック位置nにおける時間方向の明るさの変化を表し、 $avgU[n]$ ,  $avgV[n]$ の時間系列は各マクロブロック位置nにおける時間方向の色の変化を示す。また、 $stdY[n]$ は、図8の波形の平均 ( $avgY[n]$ ) からのバラツキの度合いを示し、 $stdU[n]$ ,  $stdV[n]$ もそれぞれ平均 ( $avgU[n]$ ,  $avgV[n]$ ) からのバラツキの度合いを示している。したがって、これらの特徴記述子セットは、イントラフレーム系列の明るさと色に関する時間的変化を特徴づける記述子と考えることができる。波形をそのまま記述子として保持するとビデオセグメントの長さに応じて際限なく記述子の数が

50

(10)

17

増えてしまうが、以上の記述子を用いることで時間的変化の特徴を保持しつつ、ビデオセグメントの長さ依存せずに一定数の記述子によってビデオセグメントの明るさ、色に係る特徴を記述することが可能となる。

【0070】一方、インターフレームに関しては、各インターフレームごとに生成した特徴記述子セット ( $C_{Np}$ ,  $N_{Sr}$ ,  $N_{Mr}$ ,  $N_{Ir}$ ) をインターフレームの総数  $N_p$  で平均する。よって、インターフレームについてはビデオセグメントあたり4つの特徴記述子セットが生成される。

【0071】さらに、特徴記述子セットは、検索に関して利便性が高くなるよう構造化して表現する。特徴記述子セットのデータ形式としては、属性リストと同様、例えばXML文書とする場合がある。

【0072】図9に、抽出された特徴量から特徴記述子セットを生成する処理の一例を示す。図9に示すように、特徴抽出部3にてビデオビットストリームより抽出された特徴量セットは、DTDに従って特徴記述子セットであるXML文書に変換される。図9に示す特徴記述子セットには、上述した動きベクトルに関する特徴記述子セットと輝度・色差信号の直流成分に関する特徴記述子セットが含まれている。〈MotionInfo〉〜〈/MotionInfo〉で囲まれた情報が動きベクトルに関する特徴記述子セットであり、( $C_{Np}$ ,  $N_{Sr}$ ,  $N_{Mr}$ ,  $N_{Ir}$ ) のそれぞれがAverage, ShortRun, MediumRun, LongRunのタグで囲まれて記述されている。

【0073】また〈ColourTexture〉〜〈/ColourTexture〉で囲まれた情報が輝度・色差信号の直流成分に関する特徴記述子セットであり、その中の〈YDC〉〜〈/YDC〉で囲まれた情報が輝度信号の直流成分に関する特徴記述子セットであり、( $avgY[n]$ ,  $stdY[n]$ ) のそれぞれがAverage, Stdのタグで囲まれて記述されている。また同様に〈UDC〉〜〈/UDC〉または〈VDC〉〜〈/VDC〉で囲まれた情報が色差信号の直流成分に関する特徴記述子セットであり、それぞれの中でAverage, Stdのタグで囲まれて記述されている情報が ( $avgU[n]$ ,  $stdU[n]$ ) , ( $avgV[n]$ ,  $stdV[n]$ ) である。

【0074】また図9に示す特徴記述子セットには、上述した動きベクトルに関する特徴記述子セットおよび輝度・色差信号の直流成分に関する特徴記述子セットの他に、メディアに関する補助情報が含まれている。図9において〈MediaInfo〉〜〈/MediaInfo〉で囲まれた情報がそれであり、MPEG-1やJPEG等の画像フォーマット (Format)、CD、DVD等の蓄積媒体の種類 (Medium)、その他、カラー画像かモノクロ画像かといった色に関する情報や、画像サイズ、その他画像の表示に必要な端末能力の情報等の補助情報が含まれている。これらの情報は、検索した画像をユーザが再生、表示する際に必要であると同時に検索条件として指定することも可能である。また、補助情報には、画像が蓄積されているアドレスを示

18

す情報が含まれている。図9の例では〈Locator〉〜〈/Locator〉で囲まれた情報がそれであり、URLで指定している。

【0075】尚、図9に示す例以外の特徴記述子として、画像の意味的な特徴を記述した特徴記述子セット等がある。これは例えば、スポーツ映像に写っている選手の所属しているチーム名や、映像に写っている人物の関係等を表現する特徴記述子である。これらの情報はホームビデオやデジタル写真などを家庭内サーバに蓄積する例を想定する場合には、被写体を持つ意味的な特徴をユーザがテキスト情報として直接入力したり、またテレビ放送を家庭内サーバに蓄積する例を想定する場合には番組情報より自動的に抽出することができる。

【0076】さらに別の特徴記述子セットとして、インターフレームの特徴記述子セット ( $C_{Np}$ ,  $N_{Sr}$ ,  $N_{Mr}$ ,  $N_{Ir}$ ) に対して、ビデオセグメント内の平均だけでなく、その時間的変化の複雑さ、すなわち動きの推移に関する複雑さを表す指標として、前記イントラフレームの特徴記述子セットと同じように、ビデオセグメント内の分散 (標準偏差) を算出して特徴記述子セットに加えてもよい。これにより、動きの推移に関しても時間的変化の特徴量を加えることが可能である。

【0077】さらに、イントラフレームの特徴記述子セットについて、インターフレームの特徴記述子セット生成と同様の処理を行って特徴記述子の数を削減することも考えられる。たとえば、各イントラフレームについて下式6の量 ( $C_Y, N_I$ ,  $C_U, N_I$ ,  $C_V, N_I$ ) を求める。これらはそれぞれ、 $N_I$  番目のイントラフレームに関する輝度の画面内平均、色差の画面内平均に相当する。

【0078】

【数6】

$$C_{Y,N_I} = \frac{1}{N_{mb}} \sum_{n=1}^{N_{mb}} DC_{Y,N_I}[n]$$

$$C_{U,N_I} = \frac{1}{N_{mb}} \sum_{n=1}^{N_{mb}} DC_{U,N_I}[n]$$

$$C_{V,N_I} = \frac{1}{N_{mb}} \sum_{n=1}^{N_{mb}} DC_{V,N_I}[n]$$

【0079】これらの値をインターフレームにおける  $C_{Np}$  と同様に用いて、画面内で輝度・色差に関して

( $N_{Sr}$ ,  $N_{Mr}$ ,  $N_{Ir}$ ) に該当する値を得ることが可能である。これらの値のセットは、いわば輝度・色差の空間内の分布の複雑さを記述する量であるといえる。これらを各イントラフレームの特徴記述子セットとし、ビデオセグメント内で平均、分散 (標準偏差) をとることにより、ビデオセグメントトータルの記述子数を削減した記述も可能である。逆に、インターフレームの特徴記述子セットを、イントラフレームの特徴記述子のように、より詳細なマクロブロックの精度で時間的変化を含めて記述するようにしてもよい。

【0080】また、本実施の形態1で述べたイントラフレームのマクロブロック単位の特徴記述子セットにおい

(11)

19

て、階層的な検索を実現するために、特徴記述子セットを、図10に示すように上位レイヤから下位レイヤの階層的に生成して記述することも可能である。これにより、画面全体を対象とする大局的な輝度・色差の変化から、画面内のマクロブロックを単位とする局所的な輝度・色差の変化までを一括して記述可能であり、ニーズに応じた検索処理を実現可能である。さらに、階層数を別途指定すれば、特徴記述子セットとして目的にあった階層数分だけのデータを保持すればよいので、特徴記述子セットのデータ量も適宜調整することができる。

【0081】さらに、本実施の形態1におけるイントラフレームの特徴記述子セットの数は、マクロブロックの数に依存しているため、マクロブロック数の異なるビデオセグメントとは厳密な適合判定処理ができないが、この問題に対処するためには、検索元と検索先のそれぞれの画像データのマクロブロック数に応じて、適宜いずれかの特徴記述子セットを補間または間引いてマクロブロック数を合わせてからマッチングをとるようにすればよい。このようなケースは、様々な画像フォーマットのビデオコンテンツが混在するデータベースに対して検索をかける処理に該当するが、実質的なアプリケーション（例えば、デジタル放送など）では導入設備などの都合上、画像フォーマットがほぼ統一されることが多く、事例としては多くないと考えられる。なお、本実施の形態においては、特徴記述子セットの記述にXMLを用いたが、属性リストのときと同様に、特徴記述子セットも、XMLに限らず、予め定義されたデータ構造に基づいて記述されていけばよい。

【0082】以上のようにして、ステップS5の処理により特徴記述子セット生成部4によって特徴記述子セットが生成されると、その生成された特徴記述子セットは、画像データとともに画像情報蓄積部5に蓄積される（ステップS6）。なお、特徴記述子セットと画像データとは、異なるデータベースに蓄積されることも可能である。この場合、特徴記述子セットに画像データの記憶位置を示すアドレスが含まれていけばよい。図9に示す例においては、特徴記述子セットにLocatorという要素が含まれており、この要素の値にURLを設定している。

#### 【0083】(2) 画像検索部8の処理

次に画像検索部8の処理について説明する。図11に、画像検索部8の処理を示すフローチャートを示す。画像検索の処理は主に2つの処理からなる。属性リストに基づく検索と、特徴記述子セットによる特徴類似検索である。

【0084】先に、属性リストに基づく検索処理について詳述する。まず、ユーザインタフェース部9を介してユーザから検索条件が入力する（ステップS18）。ここでは、例えば検索しようとする画像のカテゴリや、画像が作成された日時、画像の作成者等のテキスト情報を入力する。ユーザから入力されたテキスト情報は、検索処

20

理部10に渡され、検索処理部10では、属性リスト蓄積部7に蓄積された属性リストと適合判定処理を行う

（ステップS19）。例えば、ユーザからの検索条件として、画像の作成日時と画像の作成者が指定された場合を考える。図3の例では、Date、Time、Authorのタグを探索し、タグに囲まれる値とマッチングをとればよい。

【0085】属性リストとの適合判定処理の結果、Date、Time、Authorのすべての要素が一致したビデオメールの属性情報があった場合は（ステップS19“Y”）、適合結果であるDate、Time、Authorのすべての要素が一致したビデオメールの属性情報を属性リストから取り出し、表示部11へ渡す（ステップS20）。このように、画像データの属性情報であるテキスト情報が構造化されて記述されているため、データ構造の定義に基づき、検索に必要な要素のみを取り出し、容易にマッチングをとることができる。

【0086】以上のステップS20、21の処理が属性リスト蓄積部7に蓄積されている属性リストの最後まで繰り返され（ステップS21“N”）、属性リストの最後まで終了すると（ステップS21“Y”）、ステップS20の処理によりユーザからの検索条件に適合した属性情報はすべて表示部11に渡されることになる。

【0087】蓄積されている属性リストの最後まで適合結果の転送処理が終了すると（ステップS21“Y”）、表示部11では、ユーザからの検索条件に適合した属性情報の一覧を表示する（ステップS22）。図3の例では、Date、Time、Authorの値がユーザから入力された値と一致した属性情報の一覧をテキスト情報として表示する。このとき、図3の例のようなXML文書で属性データが記述されている場合には、表示部ではXML文書をブラウザ上で表示可能なHTML形式に変換し表示すればよい。また、その際に属性情報にハイパーリンクを割り付けておくことで、次段のユーザアクセスを容易に行えるようにできる。画像データが映像データであって所定の表示面積がある場合は、属性情報と一緒に映像データのキーフレームを表示させ、キーフレームデータから映像データへのリンクを張っておけば、ユーザは直観的に映像データへのアクセスを行えるようになる。

【0088】以上述べたように、属性リストが予め決められたデータ形式により構造化されているため、ユーザ側ではデータ形式を定義したファイル（例えばDTD）により、属性リストのデータ構造を解析し、表示可能なデータ形式に容易に変更することができる。

【0089】次に、表示された属性リストの中に所望の画像の属性情報がある場合、ユーザは、インターフェース部9を介して再生指示を入力する（ステップS23）と同時に、再生したい画像の属性情報を選択する（ステップS24）。すると、検索処理部10では、ユーザによって選択された画像の属性情報に含まれるアドレス（URL）から画像データを表示部11へ転送し、表示部11

(12)

21

はその画像データの再生／表示を行う（ステップS25）。

【0090】一方、表示部11に表示された属性リストの一覧の中に所望の画像の属性情報に一致はしないが、所望の画像の属性情報に近い属性情報がある場合、あるいは属性フレームとともに表示されているキーフレームが所望の画像に類似している場合は、ユーザは、インターフェース部9を介して類似検索指示を入力する（ステップS23）と同時に、類似検索に用いる所望の画像の属性情報に近い属性情報を選択する（ステップS26）ことにより、所望の画像と類似する画像を画像情報蓄積部5にて検索することができる。

【0091】つまり、ユーザによりインターフェース部9を介して類似検索指示が入力し（ステップS23）、続いて類似検索に用いる属性情報が選択されると（ステップS26）、検索処理部10は、特徴記述子セットによる特徴類似検索を行ない、ステップS26により選択された画像の特徴記述子セットを画像情報蓄積部5より抽出して、画像情報蓄積部5に蓄積されている他の画像の特徴記述子セットとの適合判定処理を行う（ステップS27）。尚、以下の説明では、上述したように特徴量が動きベクトルや輝度・色差信号の直流成分に関する特徴記述子セットで記述されている場合を例にとり説明する。

【0092】図12は、図11におけるステップS27の適合判定処理の一例を詳細に示すフローチャートである。まず、ビデオセグメント内のインターフレームの特徴記述子である動きベクトルに関する4つの特徴記述子  $C_{Np}$ 、 $N_{Sr}$ 、 $N_{mr}$ 、 $N_{lr}$ と、イントラフレームの特徴記述子である輝度・色差成分の直流成分に関する  $N_{mb}$ 個の特徴記述子  $avgY[n]$ 、 $avgU[n]$ 、 $avgV[n]$ 、 $stdY[n]$ 、 $stdU[n]$ 、 $stdV[n]$ のそれぞれについて、ステップS26により選択された画像の特徴記述子セットの各特徴記述子の値に基づいて閾値により規定した当該各特長記述子の所定の範囲内にあるか否かを比較し（ステップS27-1～S27-11）、すべての特徴記述子の値が所定の範囲内の値をとる場合には適合であると判定し、それ以外の場合は不適合とする。

【0093】そして、適合であると判定された特徴記述子を持つ画像のアドレス情報（URL）は、表示部11へ転送される（ステップS28）。以上のようにして、特徴記述子セットリスト中の全ての特徴記述子セットについて適合判定処理が終了するまで（ステップS29 “Y”）、同様に残りの特徴記述子セットとの適合判定処理を行う。

【0094】そして、特徴記述子セットリスト中の全ての特徴記述子セットについて適合判定処理が終了した場合（ステップS29 “N”）、表示部11は、適合であると判定された特徴記述子を持つ画像のアドレス情報に基づき画像情報蓄積部5から画像を取り出し、その画像を再生／表示する（ステップS22）。なお、動画画像の場合

22

には、適合であると判定された特徴記述子を持つ動画画像のキーフレームのアドレス情報を表示部11へ転送し、表示部11はそのキーフレームを再生／表示する。そして、ユーザは、表示部11に表示されたキーフレームの一覧から表示したい動画画像をインターフェース部9により選択し（ステップS24）、動画画像を再生／表示する（ステップS25）。尚、再度、類似検索したい画像を選択し（ステップS26）、上述したように類似検索を行う（ステップS27）ことも可能である。

【0095】従って、本実施の形態1の画像検索システムによれば、属性リストを参照しての属性情報に基づく検索を行った後に、特徴記述子に基づく検索を行うようにしたので、効率よく検索を行うことができる。

【0096】なお、ステップS27の適合判定処理については、他にも様々な方法が考えられる。つまり、上述した図12の例では、すべての特徴記述子について均一な重みをつけて閾値との比較処理を行っているが、例えば動きに重点をおいて検索したい場合は、インターフレームの特徴記述子の比較処理の結果に重みをかけて最終的な結果を導くよう構成することもできる。さらに、図12の例では一つ一つの特徴記述子についての比較結果をカスケード接続して最終的な結果を得ているが、全部の特徴記述子の比較処理を適切な正規化処理のもとで結合して、一括判定させるよう構成することもできる。これらの適合判定処理は、特徴記述子のデータを検索目的に応じてもっとも効率よく検索に利用できるよう調整することができる。

【0097】また、複数種類の特徴記述子を用いた検索では、あらかじめアプリケーション（検索システム、データベースなど）のタイプに応じて、適合判定処理（検索式）を記述子の形態で保持しておくことも考えられる。判定処理自体を標準的な記法に従って記述することにより、検索に使用する複数の特徴記述子をいかに検索に利用するかをアプリケーションに依存することなく記述することができるので、様々なタイプのアプリケーションを統一的に実装することができる。また、例えば、インターネットで接続され、まったく同様の特徴記述子を検索に利用する別のデータベース上のデータを共通の適合判定処理によって検索処理することも可能となる。適合判定処理の記述例としては、例えば、以下の（1）～（5）に示すようなものが考えられる。

【0098】（1）複数の特徴記述子をどのように適合判定に利用するかを記述する。例えば、一括検索か、カスケード検索か等である。

（2）一括検索処理を行うにあたっての各特徴記述子の重要度に応じた重み係数を与える。

（3）一括検索を行うにあたって、各特徴記述子の正規化方法を与える。

（4）記述子を検索に使用する順序を与える。

（5）カスケード接続による検索の場合に、各特徴記述

(13)

23

子によって得られる検索候補の数を与える。

(6) 各特徴記述子の記述の精度（どれくらい正確に特徴を記述しているか）を示す値を与える。

(7) 各特徴記述子のマッチングによって得られる検索候補を、それぞれANDして出力するか、ORして出力するか判断基準を与える

【0099】これらの適合判定処理手順の記述子は、アプリケーション固有の固定的な記述であってもよいし、アプリケーションが許容する範囲においてユーザによるカスタマイズが可能となるよう構成することもできる。例えば、ユーザは検索利用可能な特徴記述子の種類をシステムから知らされ、それらの性質に応じて自らの好みを反映した独自の適合判定処理を記述子として更新できるようにすれば、ユーザにとってより柔軟な検索処理を行うことができるようになる。このような検索システムのカスタマイズはある程度アプリケーション独自に実装することができるが、共通の標準的な記述フォーマットを用意することにより、広範なシステムの実装が容易になるほか、検索に利用可能なデータを異なるアプリケーション間で共有することも可能である。

【0100】また本システムの応用例として、本システムを監視システムに適用することが考えられる。例えば、監視カメラにより侵入者の監視を行う監視システムの場合、画像の動きの特徴量により、侵入者の検出が可能であるから、画像の動きの特徴量を記述したメタデータを監視画像ビットストリームとともにデータベースに登録し、属性リストには、監視画像が記録された日時等の情報を登録する。ユーザは記録された監視画像を再生する際に、検索キーとして、日時等のテキスト情報を入力し、さらに「侵入者」というキーワードを検索キーとして入力する。登録されたリストから検索画像候補を選択し、次に検索画像候補に付随するメタデータを検索し、侵入者のあったと思われる画像を表示する。ユーザは表示された画像から見たい画像を選択して再生することができる。さらに、一旦検索して得られた映像に対し、正確な記録日や記録時刻が不明な類似映像をさらに詳細に蓄積映像中から検索するような要求に対しても、特徴記述子セットを用いて映像間のマッチングをとることによりユーザに対して詳細な検索結果を与えることが可能となる。

【0101】また、本システムの別の応用例として、携帯端末等に蓄積されたビデオメールや、ホームサーバに蓄積された放送番組から、ユーザが見たいメールあるいは放送番組を検索して再生するシステムに適用することが考えらる。属性リストは、ホームサーバに番組が録画された日時や番組情報から抽出された番組のタイトル、出演者等の情報から構成される。ユーザは見たい番組の情報を入力することにより検索を行うが、検索結果に所望の情報が得られない時には映像に付加された特徴記述子セットによる類似検索を行うこともできる。これによ

24

り正確な番組の情報を覚えていなくても映像に対する印象により検索することも可能となる。

【0102】さらに、以上述べた特徴抽出部3、特徴記述子セット生成部4の動作や抽出される特徴量、生成される特徴記述子セットについては、他にも様々な例がある。例えば、図12に示すように、特徴抽出部3がビデオ圧縮符号化部12により圧縮符号化される前の非圧縮状態の画像データから特徴量を抽出して、それらを特徴記述子セット生成部4に受け渡す場合も考えられる。このような構成されたシステムの応用例としては、例えば、長時間の監視映像を圧縮しながら蓄積するシステムにおいて、カメラからのデジタイズされた入力映像に対して直接移動物体や侵入物などの特徴（色、形状、サイズ、動きの大きさ・方向など）を検出して特徴量として抽出し、特徴量記述子セット生成部4において特徴記述子セットを生成した後、ビデオ圧縮符号化部12において圧縮符号化されるビデオセグメントに付加して蓄積するという運用形態が考えられる。この例では、画像データとしての詳細な特徴を保持した非圧縮映像について特徴量を抽出するので、圧縮データから特徴量を抽出する場合に比べてより詳細な特徴記述子（被写体の種類や軌跡の記述など）を生成できるという利点がある。一方、詳細な特徴記述子の生成には複雑な演算処理系が必要となり、特徴抽出部3、特徴記述子セット生成部4には、より高い演算性能が要求されることになる。

【0103】このような系は、監視だけでなく、コストよりも品質・信頼性を重視する放送業務用途のビデオカメラにも適用可能である。その場合は、本実施の形態に述べたシステム全体が放送番組制作用の映像ライブラリシステムとして機能する。ビデオカメラから取得された映像の特徴記述子セットが記録された映像と一緒にデータベースに登録されることにより、映像自体の持つ類似性をキーとして効率的な素材検索を行うことが可能となる。また、パーソナルユースのビデオカメラにおいても、圧縮符号化の前に簡易な前処理機能を持たせることで、例えば人物の顔画像に関する特徴記述子を生成して、対応する映像と一緒に媒体（ビデオテープでもPCもしくはホームサーバ上のハードディスクでもよい）に記録できるようにしておけば、膨大な未整理映像の中から特定の顔に該当する映像を特徴記述子セットのマッチングによって検索することが可能となる。

【0104】以上のいずれのアプリケーションにおいても、属性リストを用いることにより、複数の媒体やデータベースを記録映像の属性の観点から一元管理することができ、さらに検索候補となる映像の追加や削除などを属性リストのレベルで簡単に行うことができるようになるため、検索処理だけでなく、システムの構築自体を統一かつ簡易に行うことができる。

【0105】実施の形態2. 本実施の形態2では、特に動画像データに関して、動きや明るさ、色など、複数の



(14)

25

種類の特徴記述子を組み合わせる特徴記述子セットを構成し、これらを検索のキーとして用いて、ユーザの意図する検索方針の反映等を可能にした画像検索システムおよび画像検索方法の具体例について説明する。尚、上記実施の形態1では、検索処理部10をサーバ側に設けて説明したが、本実施の形態2の画像検索システムでは、検索処理部10をサーバ側とクライアント側とに分けて、サーバ側とクライアント側とを2つのネットワークA、Bを介し接続したことを特徴の一つとするものである。

【0106】図15は、本実施の形態2の画像検索システムにおけるサーバ側の特徴部分の構成を示すブロック図である。つまり、図15は、図1における特徴記述子セット生成部4の詳細な構成を示すと共に、検索処理部10を2つの検索処理部10A、10Bに分割したうちのサーバ側の検索処理部10Aの詳細ブロックを示し、かつ検索結果を再生するために必要なビデオ再生サーバ218とを示したものである。

【0107】図15において、201は圧縮ビデオデータ、202は特徴記述子抽出部、203は特徴記述子データ、204は特徴記述子信頼度算出部、205は特徴記述子信頼度、206は特徴記述子検索使用順序決定部、207は特徴記述子検索使用順序(サーチプライオリティ)データ、208は圧縮ビデオデータ格納位置情報、209は検索用記述データ生成部である。

【0108】また、210は検索用記述データファイル、211は検索用記述データ解析部、212は解析された特徴記述子データ等、213は適合判定処理部、214は適合判定処理結果、215は検索結果符号化・伝送部、216は検索要求情報、217は検索結果情報、218はビデオ再生サーバ、219は再生要求情報、220は再生制御部、221は再生コンテンツ特定情報、222は再生するビデオコンテンツデータ、223はビデオデータ送信部、224は配信ビデオデータである。

【0109】図16は、本実施の形態2の画像検索システムにおける主にクライアント側の特徴部分の構成を示すブロック図である。つまり、図16は、図1における検索処理部10を2つの検索処理部10A、10Bに分割したうちのクライアント側の検索処理部10Bの詳細ブロックを示すと共に、検索結果をユーザのために再生するビデオ復号・再生部、ユーザとのインタラクション、再生ビデオの表示などのために設けられたユーザインタフェース等を示す。

【0110】図16において、225はネットワークA、227は検索要求情報符号化・送信部、228は検索結果受信・表示部、229はユーザインタフェース部、230はユーザからの検索要求、231はユーザインタフェース部229に表示される検索結果、232はユーザ、233は検索結果再生要求、234は検索動画再生、235はビデオ再生制御部、237はネットワー

26

クB、239はビデオ復号・再生部である。

【0111】尚、図16に示すクライアント側の検索処理部10B、ビデオ再生制御部235、およびビデオ復号・再生部239と、図15に示すサーバ側の検索処理部10Aおよびビデオ再生サーバ218との間では、ネットワークA225、B237を介して、検索要求情報216、検索結果情報217、再生要求情報219および配信ビデオデータ224をやり取りする。ここで、ネットワークA225は、例えば、データ伝送の高速性よりも信頼性に重点を置いたネットワークを意味し、たとえば、TCP/IPなどのコネクションを意味する。また、ネットワークB237は、例えば、データ伝送の信頼性よりもリアルタイム伝送のための高速性を重視したネットワークを意味し、たとえば、RTP/UDP/IPなどのコネクションを意味する。

【0112】また、図16におけるユーザインタフェース部229は、図1におけるクライアントの部分に該当し、図1におけるユーザインタフェース部9と表示部11を併せ持つクライアントアプリケーションを示すものとする。

【0113】また、上述したように、図15に示す構成は主としてサーバ側に設置されるコンポーネント、図16に示す構成は主としてクライアント側に設置されるコンポーネントを意図している。これらの図からわかるように、本システムでは、実際の検索適合判定処理などの検索処理自体は検索処理部10Aによりサーバ側で行われることを想定し、クライアント側は検索処理に対するユーザの趣向などの要求情報を所定のルールにしたがって符号化してサーバ側に送信することにより、サーバがそれらの情報を解釈して検索処理に反映させることを想定する。

【0114】このようなシステムとすることにより、多数の検索用のビデオ記述データを格納するファイルを、ネットワークを介してクライアント側に送信する必要がなくなり、ネットワーク使用効率を高めることができる。また、通常の検索結果の提示形態としては、いくつかの検索候補をユーザに提示する形が一般的であるが、検索結果情報として検索候補となったビデオデータすべてをクライアント側に伝送することなく、リアルタイムビデオ再生系と組合せることによって、ユーザが真に必要な検索結果だけをクライアント側に伝送することができる。これもトータルのネットワーク使用効率を高める効果が得られる。

【0115】本システムの検索処理の特徴として、複数のビデオ特徴記述子に対して、それぞれ記述の信頼性を示す数値を算出し、その値に応じて、検索に使用する順番を定め、それらの値を記述データの一部としてXMLファイルなどの検索用記述データに格納する点が挙げられる。また、この、検索に使用する順序に関する記述子は、検索に関するユーザの趣向などを反映するため、ク



(15)

27

ライアントからのユーザ要求に応じて更新可能なデータとすることによって、複数のビデオ特徴記述子の組合せ検索のパラメータを与える記述子として機能させる。

【0116】また、各特徴記述子の信頼性を示す数値は、各特徴記述子を検索に用いるステップでの検索結果絞り込み候補数などのシステムパラメータの決定に利用する。以下、特徴記述子を検索に使用する順序の記述子を「サーチプライオリティ」、各特徴記述子の信頼性を示す数値を「信頼度」と呼ぶ。両者の詳細な意味、使用方法は後述する。

【0117】図17に、本システムにおける検索用記述データ生成および検索処理のフローチャートを示す。以下、図15～17に基づいて、本実施の形態2におけるシステムの詳細な動作を説明する。

#### 【0118】(1) 検索用記述データファイル210の生成 (S30～S33)

はじめに、検索用記述データを付加しようとする入力画像データとしての圧縮ビデオデータ201に対し、以下に述べる処理を行うことにより、検索用記述データファイル210の生成を行う。以下では、検索用記述データを生成する対象となる圧縮ビデオデータ201をビデオセグメントと呼ぶ。ビデオセグメントは、実施の形態1にも述べたように、任意の数のフレーム群から構成される単位ビデオデータを示す。したがって、ビデオセグメントは、たとえば1本のテレビプログラムやビデオソフトを示す場合もあるし（以下、このような場合を「ビデオプログラム」と呼ぶ）、テレビプログラム中の特定の意味内容に応じて分類されるシーン（以下、「ビデオシーン」と呼ぶ）や、シーンを構成する個々のカメラショット（以下、「ビデオショット」と呼ぶ）などを示す場合もある。これらはフレーム数の観点から言えば、異なるフレーム数を持つビデオセグメントである。

#### 【0119】(1-1) 検索用記述データの構成

図18に、本実施の形態2における検索用記述データファイル210を構成する検索用記述データの構造の記述例を示す。具体的には、検索用記述データは、ビデオプログラム240を最上位階層とし、ビデオプログラム240を構成するビデオシーン（図では、「シーン」と略す）243群を次の階層とし、さらに各シーン243を構成するビデオショット（図では、「ショット」と略す）244群を最下位階層として構成される。尚、図上では、ビデオセグメント（ビデオプログラム240、ビデオシーン243、ビデオショット244）、サーチプライオリティ241、および信頼度242は「□」で示している一方、各ビデオセグメント、特にビデオショットを記述する特徴記述子245～249群を「○」で示している。本システムでは、ビデオショット244に関する以下の特徴記述子245～249のセットを生成する。

【0120】・動き量245

28

ビデオセグメント中の動きベクトルの大きさに関する特徴記述子である。

・動き方向246

ビデオセグメント中の動きベクトルの方向に関する特徴記述子である。

・動き分布247

ビデオセグメント中の動きベクトルの画面内分布に関する特徴記述子である。

・明るさ248

10 ビデオセグメント中の平均的な明るさに関する特徴記述子である。

・代表色249

ビデオセグメント中の代表色に関する特徴記述子である。

【0121】尚、各階層240～244および特徴記述子245～249には、後述する信頼度242が設定される。

【0122】また、サーチプライオリティ241については、最上位階層のビデオプログラム240に対して120 2つ付加する。

【0123】各特徴記述子に代表される特徴量は、その検索使用順序が変わることによって異なった結果を返す。この例では、多くの初期データベースの段階から、動き量的によく適合するビデオセグメントがはじめに絞り込まれるので、最終的にユーザに提示される検索結果には、確実に動き量の観点で類似度の高い結果が得られることが期待される。こういった情報をビデオセグメントに対して付与することにより、ユーザは常にサーチプライオリティ241の情報に基づいた検索結果を得ることができ、さらにユーザがこの情報を動的に変更することを可能にすることにより、ユーザの意図に応じた検索結果を柔軟に引き出すことが期待できる。

【0124】また、本実施の形態2では、ビデオプログラム240中のある特定のビデオショット244に対し、類似する他のビデオショット244を同一ビデオプログラム240中から検索する処理を想定するため、サーチプライオリティ241は、ビデオプログラム240に対して1つ付加することとし、ビデオシーン243やビデオショット244の階層に相当するビデオセグメントには付与しないものとする。

40 【0125】次に、各特徴記述子に設定される信頼度242について説明する。信頼度242とは、たとえば動き量に関する特徴記述子を例にとると、その記述子がビデオコンテンツの動き量をどれだけ正確に記述できているかを客観的な数値として表現するものである。数値化の方法については、各々の特徴記述子の生成に関する説明の項目で詳述する。本実施の形態2では、信頼度242は、ビデオショット244の個々の特徴記述子245～249等に対して算出および付与される信頼度24250 Cと、各ビデオシーン243に対し付与される各ビデオ

(16)

29

シーン243を構成する全ビデオショット244の信頼度の値の平均値である信頼度242Bと、ビデオプログラム240に対し付与される当該ビデオプログラム240を構成する全ビデオシーン243の信頼度の値の平均値である信頼度242Aとがある。検索用記述データを構成する各要素に付与された以上の信頼度242は、主にサーチプライオリティ241の初期設定と、検索処理段階でのパラメータ設定に利用する。詳細は後述する。

【0126】(1-2) 検索用記述データ生成処理  
(1-2-1) 特徴記述子の生成処理 (S30)

個々の特徴記述子の生成は、まず、特徴抽出部3が特徴記述子の生成に必要な動きベクトルやDC成分値などの画像特徴量を圧縮ビデオデータ201中から抽出し、これらに基づいて特徴記述子生成部202が各特徴記述子の生成を行う (S30)。以下、各特徴記述子の生成方法の詳細について述べる。

【0127】(a) 動き量に関する特徴記述子245  
ビデオセグメント中のk番目のインターフレーム (P-frame) に対し、下式に従い、フレーム内のすべてのマクロブロック ( $1 \leq n \leq N$ ;  $N$ はフレーム内マクロブロック総数) の動きベクトル  $MV_k[n]$  の大きさ  $C_k[n]$  を抽出する。

【0128】

【数7】

$$C_k[n] = \sqrt{x_n^2 + y_n^2}$$

$x_n$ : 動きベクトル  $MV_k[n]$  の水平成分  
 $y_n$ : 動きベクトル  $MV_k[n]$  の垂直成分

【0129】次いで、1フレーム分蓄積された  $C_k[n]$  について、以下の平均値および分散を求める。

【0130】

【数8】

$$C_k^{avg} = \frac{1}{N} \sum_{n=1}^N C_k[n]$$

【0131】

【数9】

$SR_k$ : フレーム水平画素数の1/3以下の長さをもつゼロラン (short run) の数

$MR_k$ : フレーム水平画素数の1/3以上2/3の長さをもつゼロラン (middle run) の数

$LR_k$ : フレーム水平画素数の2/3以上の長さをもつゼロラン (long run) の数

【0141】尚、実施の形態1でも説明したが、図7からわかるように、short runが多いフレームでは動きの複雑度が大きく、広い範囲に動きが分布していることを示し、long runが多いフレームでは動きが局在化されていることを示す。

【0142】そして、さらに、ビデオセグメントあたりの値を得るため、ビデオセグメント内のインターフレームの総数で平均をとる。

【0143】(d) 明るさに関する特徴記述子248  
イントラフレームの輝度信号成分に含まれる各マクロブロックの直流 (DC) 成分のヒストグラムを用いる。DC成分

30

$$\sigma_k^2 = \frac{1}{N} \sum_{n=1}^N (C_k[n] - C_k^{avg})^2$$

【0132】以上の結果、インターフレームkの動き量に関する特徴記述子のデータセット ( $C_k^{avg}$ ,  $\sigma_k^2$ ) が得られる。さらに、ビデオセグメントあたりの値を得るため、ビデオセグメント内のインターフレームの総数で平均をとる。

【0133】(b) 動き方向に関する特徴記述子246  
各インターフレームにおいて、動きベクトル  $MV_k[n]$  の角度  $A_k[n]$  を下式で求める。角度はDegree表記とする。

【0134】

【数10】

$$A_k[n] = \tan^{-1}(y_n / x_n)$$

【0135】一方、イントラマクロブロックでは、 $A_k[n]=0$ とする。1フレーム分蓄積された  $A_k[n]$  について、下式により平均をとる。さらに、ビデオセグメントあたりの値を得るため、ビデオセグメント内のインターフレームの総数で平均をとる。

【0136】

【数11】

$$A_k^{avg} = \frac{1}{N} \sum_{n=1}^N A_k[n]$$

【0137】(c) 動き分布に関する特徴記述子247

(a) においてインターフレームkについて求めた各動きベクトル  $C_k[n]$  に対して、以下の閾値処理を施す。

【0138】

【数12】

$$C_k'[n] = \begin{cases} C_k[n], & \text{if } C_k[n] \geq C_k^{avg} \\ 0, & \text{if } C_k[n] < C_k^{avg} \end{cases}$$

【0139】こうして得られた新しい動きベクトル分布に対し、 $C_k'[n]$  をラスタスキャン順にスキャンし、以下の動き分布を特徴づける量を求める。

【0140】

は0~255の範囲の値とし、その範囲を16個のサンプルへ量子化し、各サンプルに対してヒストグラムを求める。イントラフレームごとに求めたヒストグラムを各サンプルごとにビデオセグメント中の全イントラフレーム数で平均する。最後に、下式により、全イントラフレーム (N枚) の平均をとる。

【0144】

【数13】

$$H_{avg}(j) = \frac{1}{N} \sum_{i=0}^N H_i(j), \quad j=0, K, 15$$

【0145】(e) 代表色に関する特徴記述子249

(17)

31

イントラフレームの輝度および色差信号成分に含まれる各マクロブロックの直流(DC)成分のヒストグラムを用いる。DC成分は0～255の範囲の値とし、その範囲を16個のサンプルへ量子化し、各サンプルに対してヒストグラムを求める。イントラフレームごとに求めたヒストグラムを各サンプルごとにビデオセグメント中の全イントラフレーム数で平均する。最後に、下式により、全イントラフレーム(N枚)の平均をとる。

【0146】

【数14】

$$H_{avg}^k(j) = \frac{1}{N} \sum_{i=0}^N H_i^k(j), \quad k: \text{色空間}, j=0, K, 15$$

【0147】(1-2-2) 信頼度の算出処理(S31)

次いで、特徴記述子信頼度算出部204において、S30で求めた各特徴記述子245～249に対する特徴記述子信頼度242(図15における特徴記述子信頼度205と等価)を算出する(S31)。各特徴記述子ごとの算出方法を以下に述べる。

【0148】(a) 動き量245に関する特徴記述子

ビデオセグメント内の全インターフレームについて( $C_k$  avg,  $\sigma_k^2$ )の抽出を行った後、個々の要素に対して、最終的に得られるビデオセグメントとしての値と各フレームの値との差分絶対値をとり、ビデオセグメント中の最大値を0、最小値を100として正規化した各フレームの差分絶対値を、全フレームに渡って加算平均することによって算出する。これはフレームの記述精度を基準とするビデオセグメントの記述精度を示す値となる。特徴記述子としての信頼度は、( $C_k$  avg,  $\sigma_k^2$ )に対する各信頼度の平均値とする。

【0149】(b) 動き方向246に関する特徴記述子

ビデオセグメント内の全インターフレームについて $A_k$  avgの抽出を行った後、最終的に得られるビデオセグメントとしての値と各フレームの値との差分絶対値をとり、ビデオセグメント中の最大値を0、最小値を100として正規化した各フレームの差分絶対値を、全フレームに渡って加算平均することによって算出する。これはフレームの記述精度に対するビデオセグメントの記述精度を示す値となる。

【0150】(c) 動き分布247に関する特徴記述子

ビデオセグメント内の全インターフレームについて( $S R_k$ ,  $M R_k$ ,  $L R_k$ )の抽出を行った後、最終的に得られるビデオセグメントとしての値と各フレームの値との差分絶対値をとり、ビデオセグメント中の最大値を0、最小値を100として正規化した各フレームの差分絶対値を、全フレームに渡って加算平均することによって算出する。これはフレームの記述精度に対するビデオセグメントの記述精度を示す値となる。特徴記述子としての信頼度は、( $S R_k$ ,  $M R_k$ ,  $L R_k$ )に対する各信頼度の平均値とする。

【0151】(d) 明るさ248に関する特徴記述子

32

各ヒストグラムサンプルごとに、最終的に得られるビデオセグメントとしての値と各フレームの値との間の差分絶対値をとり、ビデオセグメント中の最大値を0、最小値を100として正規化した各フレームの差分絶対値を、全フレームに渡って加算平均することによって算出する。これはフレームの記述精度に対するビデオセグメントの記述精度を示す値となる。特徴記述子としての信頼度は、各ヒストグラムサンプルに対する信頼度の平均値とする。

10 【0152】(e) 代表色249に関する特徴記述子

各色空間の各ヒストグラムサンプルごとに、最終的に得られるビデオセグメントとしてのヒストグラムと各フレームのヒストグラムとの間の差分絶対値をとり、ビデオセグメント中の最大値を0、最小値を100として正規化した各フレームの差分絶対値を、全フレームに渡って加算平均することによって算出する。これはフレームの記述精度に対するビデオセグメントの記述精度を示す値となる。特徴記述子としての信頼度は、各ヒストグラムサンプルに対する各信頼度の平均値とする。

20 【0153】(1-2-3) サーチプライオリティ241の設定処理(S32)

ビデオプログラム240中に含まれるすべてのビデオショット244に対して特徴記述子245～249および信頼度242の算出が終了した後、特徴記述子検索使用順序決定部206において、ビデオプログラム240に対するサーチプライオリティ241の初期設定を行う

(S32)。初期設定方法はシステムの設計に応じて自由に規定することが可能であるが、本実施の形態2では、ビデオプログラム240中の全特徴記述子の信頼度242Aに基づいて決定する。ビデオプログラム240の各特徴記述子の信頼度242Aを $k$ 、ビデオシーンⅢの各特徴記述子の信頼度242Bを $k(m)$ 、ビデオシーンⅢ中のビデオショットⅢの各特徴記述子の信頼度242Cを $k(m, n)$ とすると、これら $k$ 、 $k(m)$ 、 $k(m, n)$ の間には、下式の関係が成立することとなる。

【0154】

【数15】

$$\varphi_k = \frac{1}{M} \sum_{m=0}^{M-1} \varphi_k(m)$$

40 【0155】

【数16】

$$\varphi_k(m) = \frac{1}{N} \sum_{n=0}^{N-1} \varphi_k(m, n)$$

【0156】従って、 $k$ の値が大きい特徴記述子ほど、ビデオプログラム240中でより良いコンテンツ記述を与えるものと考えられるため、 $k$ 値の大きいものから順に、サーチプライオリティ241を決定する。

【0157】(1-2-4) 検索用記述データファイル210の書き出し(S33)

50 以上の処理過程を経て、最後に検索用記述データ生成部

(18)

33

209において、図18の構造に従って圧縮ビデオデータ201の時空間構造に関連付けて、ビデオプログラム、ビデオシーン、ビデオショットの全特徴記述子ならびにサーチプライオリティ、信頼度を、XMLファイルなどの形式に書き下すことにより検索用記述データファイル210を生成し、生成した検索用記述データファイル210を画像情報蓄積部5へ出力して格納する(S33)。なお、この過程では、生成された検索用記述データファイル210がどのビデオデータを記述するものかを参照できるようにするため、圧縮ビデオデータ格納位置情報208も一緒にファイル化して検索用記述データファイル210として画像情報蓄積部5へ出力することにより、ファイルを解析することにより記述対象となったビデオデータの所在を特定できるようにしている。

#### 【0158】(2) 検索処理

次に、クライアント側からユーザが指定するクエリー(検索元ビデオショット)に対して行う検索処理について詳述する。検索処理は、図15に示す検索処理部10A、および図16に示す検索処理部10Bによって実行される。

【0159】検索処理部10Aは、クライアント側から送信される検索要求情報216を解釈して、クエリーを特定し、画像情報蓄積部5に格納された検索用記述データファイル210を用いた適合判定処理によって所定の検索結果を求め、所定のルールで検索結果情報217を符号化し、クライアント側に返送する。

【0160】検索処理部10Bは、ユーザからの検索要求をユーザインタフェース部229から受け入れ、それを所定の手順に従って検索要求情報216として符号化してサーバ側の検索処理部10Aへ送信するとともに、サーバから所定のルールに従って送信された検索結果情報217を受信し、ユーザに提示できる形に変換し、ユーザインタフェース部229上に表示する処理を行う。

【0161】本実施の形態2では、図18に示すように記述されたビデオプログラム240中の特定のビデオショット244を検索元ビデオセグメント(クエリー)として、同じビデオプログラム240中からクエリーに類似するビデオショットを検索する状況を想定する。

【0162】クエリーとなるビデオショットの特定方法については、様々なケースが考えられる。例えば、ビデオプログラムを制作する側があらかじめ検索のキーとなるようなビデオショットをクエリーとして登録しておき、それをユーザに提示していずれかを指定させる方法や、ビデオプログラムを視聴するユーザが任意の場所で再生を停止し、その時点のビデオショットをクエリーとして指定するなどが考えられる。以下では、クエリーが特定された後の検索処理について詳述する。

#### 【0163】(2-1)クエリーの特定と検索用記述データの解析(S34~S36)

ユーザは、まず、ユーザインタフェース部229上に提

34

示されたクエリーリストなどから検索元となるクエリーを選定する(S34)。クエリーリストとしては、あらかじめサーバ側の画像情報蓄積部5から、クエリー候補のビデオセグメントを代表するキーフレームなどのデータをユーザインタフェース部29に提示しておく形式などを想定する。

【0164】検索処理部10Bは、検索要求情報符号化・送信部227において、ユーザから指定されたクエリーをサーバ側で特定するための情報を検索要求情報216として検索処理部10Aに送信する。これは、例えば、ビデオプログラムおよびその構成要素であるビデオショットの時系列番号であったり、適当なIDデータであればよい。また、このようなコンテンツ特定情報は、もともと検索用記述データファイル210の中にコンテンツへのリンク情報として記述されており、ユーザにクエリーリストを提示する際にコンテンツ特定情報を付加してクライアントに送信するという構成にすることもできる。

【0165】すると、検索処理部10Aでは、検索処理部10Bから送られる検索要求情報216を検索用記述データ解析部211が受け取り、検索用記述データ解析部211は、受信した検索要求情報216に基づいてクエリーとなるビデオショットに関する検索用記述データファイル210を特定し、それを解析して特徴記述子などのデータを抽出する(S35)。

【0166】また、検索用記述データ解析部211は、クエリーが属するビデオプログラムのサーチプライオリティ241と、各ビデオシーンの信頼度242Bのデータをあらかじめ抽出して、適合判定処理部213へ引き渡しておき、クエリーに関する特徴記述子などのデータを適合判定処理部213に引き渡した後、クエリーが属するビデオプログラム中の各ビデオショットに対応する検索用記述データファイル210を順次解析して、適合判定処理部213に引き渡す(S36)。

【0167】なお、検索用記述データファイル210の構成方法は任意であり、ビデオプログラム240内のすべてのビデオショット243の検索用記述データが1つのファイル内に格納されていてもよいし、個々のビデオショット244ごとにファイル分割されていてもよい。後者の場合は、ビデオプログラム240の記述データファイル210から個々のビデオショット244の記述データファイルを特定するリンク情報が含まれていればよい。

#### 【0168】(2-2) 検索ルールの決定および適合判定処理(S37~S38)

適合判定処理は、検索処理部10Aの適合判定処理部213にて行われ、その動作は大きく分けると、検索ルールの決定と、各特徴記述子による適合判定処理に分類される。

#### 【0169】(2-2-1) 検索ルールの決定

50

(19)

35

適合判定処理部213は、検索用記述データ解析部211より受け取ったビデオプログラムのサーチプライオリティ241と、ビデオシーンの信頼度242Bのデータとに基づき、検索ルールおよび検索のためのパラメータを設定する(S37)。本実施の形態2では、後述する図19に示すようにサーチプライオリティ241に基づいて各特徴記述子を検索に使用する順序を検索ルールとして決定し、その次に、信頼度に基づいて、以下の検索パラメータを決定する。

【0170】(a) 各特徴記述子による検索ステップにおける検索候補数

まず、検索結果としてユーザに返信する最終検索結果数Tを定める。これはシステムがデフォルト値としてもっていてもよいし、ユーザ要求としてユーザインタフェース部229→検索処理部10B→検索処理部10Aという順に受け渡されて設定されるようにしてもよい。検索エンジンは、各ビデオシーンについてT個以下の検索結果を求め、検索候補となった全ビデオシーン中の全ビデオショットのうち、適合判定結果の上位T個のビデオショットを最終的な結果としてユーザに返すものとする。

【0171】ビデオシーンmの検索時における検索候補数 $T_k$ を、信頼度 $k(m)$ に基づき、以下のように決定する。

【0172】

【数17】

$$T_0 = T * 5$$

$$T_k = \frac{T_0}{P_k} + [T * (1 - \frac{\varphi_k(m)}{100})]$$

【0173】ここで、 $P_k$ は、 $K(1 \leq k \leq 5)$ 番目の特徴記述子のサーチプライオリティ( $1 \leq P_k \leq 5$ )である。この設定により、各特徴記述子をサーチプライオリティの順序で検索する各ステップにおける検索候補数 $T_k$ が決定される。基本的に各検索ステップごとに検索候補数 $T_k$ が絞られていくので、検索処理を高速化することができる。場合によってはこのような検索によりユーザの真に期待するビデオショットをある検索ステップで見落としてしまう可能性もあるため、各検索ステップの検索候補\*

$$W = C^{avg}(S_q) + \sigma^2(S_q)$$

$$D(S_q, S) = \frac{W}{C^{avg}(S_q)} |C^{avg}(S_q) - C^{avg}(S)| + \frac{W}{\sigma^2(S_q)} |\sigma^2(S_q) - \sigma^2(S)|$$

(b) 動き方向に関する特徴記述子

クエリー $S_q$ 、検索先ビデオショットSの間のマッチング評価尺度 $D(S_q, S)$ を下式で定める。この $D(S_q, S)$ が小さいものほど類似度が高く、適合するものと判定する。

【0181】

【数20】

$$D(S_q, S) = |A^{avg}(S_q) - A^{avg}(S)|$$

36

\* 数は一定として、そのなかでの最終検索候補の決定に際して信頼度を利用する方法も考えられる。

【0174】(b) 閾値処理に基づく検索スキップ判定  
各ビデオショットの適合判定処理の際、下式の条件を満たすビデオショット、すなわちサーチプライオリティの最も高い特徴記述子の信頼度が極端に低いビデオショットは、検索条件に見合わないものとして適合判定処理の候補から外すものとする。信頼度は0に近いほど信頼性が低く、100に近いほど信頼性が高いとみなすので、下式のTHは適当に小さな値を設定しておけばよい。

【0175】

【数18】

$$\varphi_k(m, n) < TH; P_k = 1$$

【0176】本判定により、明らかに検索結果候補として適当でないビデオショットを適合判定処理を行う前に除外することができ、検索処理の高速化を図ることが可能である。

【0177】尚、本実施の形態2の説明では、サーチプライオリティの最も高い特徴記述子を判定に用いることとしたが、これ以外にも信頼度を用いた様々なスキップ判定方法を実現可能である。また、本実施の形態2の判定は、各ビデオショットに対する適合判定処理を行う際に、ビデオショットごとに行われる。

【0178】(2-2-2) 各特徴記述子の適合判定処理  
以下の基準に基づいて、各検索ステップにおいてサーチプライオリティ241によって定められる特徴記述子の適合判定処理を行う。(S38)

【0179】(a) 動き量に関する特徴記述子

個々のデータ間のユークリッド距離をすべて加算する。各データ要素の値域の違いを吸収するため、正規化処理を行う。クエリーを $S_q$ 、検索先ビデオショットをSとし、 $S_q$ とSとの間のマッチング評価尺度 $D(S_q, S)$ を下式で定める。この $D(S_q, S)$ が小さいものほど類似度が高く、適合するものと判定する。

【0180】

【数19】

【0182】(c) 動き分布に関する特徴記述子

クエリー $S_q$ 、検索先ビデオショットSの間のマッチング評価尺度 $D(S_q, S)$ を下式で定める。この $D(S_q, S)$ が小さいものほど類似度が高く、適合するものと判定する。

【0183】

【数21】

(20)

37

$$D(S_q, S) = |SR(S_q) - SR(S)| + |MR(S_q) - MR(S)| + |LR(S_q) - LR(S)|$$

## 【0184】(d) 明るさに関する特徴記述子

クエリー $S_q$ 、検索先ビデオショット $S$ の間のマッチング評価尺度 $D(S_q, S)$ を下式で定める。ただし、 $H_i(S)$ を $i$ 番目のサンプルのヒストグラム値とする。この $D(S_q, S)$ が小さいものほど類似度が高く、適合するものと判定する。

【0185】

【数22】

$$D(S_q, S) = \sum_{i=0}^{15} |H_i(S_q) - H_i(S)|$$

$$D(S_q, S) = \sum_{i=0}^{15} [H_i^R(S_q) - H_i^R(S)] + [H_i^G(S_q) - H_i^G(S)] + [H_i^B(S_q) - H_i^B(S)]$$

【0188】図19に、適合判定処理部211におけるサーチプライオリティ241によって定められる特徴記述子の適合判定処理の一例を示す。例えば、ビデオセグメントあたり3つの特徴記述子 $D_1 \sim D_3$ があるとする。サーチプライオリティ $P_k$  ( $k$ は自然数)は、各特徴記述子 $D_k$ が検索に使用される順序を規定する値である。例えば、 $D_1$ が動き量、 $D_2$ が動き分布、 $D_3$ が明るさに関する特徴記述子であるとする、図19の例では、適合判定処理部211は、クエリー(検索元ビデオショット)250に対し、動き量に関する特徴記述子 $D_1$ を最初に用いて適合判定処理251を行い、検索用記述データ解析部211から転送される画像情報蓄積部5に格納されていた多数のデータベース257の中から、動き量の観点で類似度が高くよく適合するビデオセグメントデータを絞り込む。その絞り込み結果252に対して、動き分布に関する特徴記述子 $D_2$ を用いて適合判定処理253を行い、動き分布の観点で類似度が高くよく適合するビデオセグメントデータを絞り込み、さらにその結果254に対して、明るさに関する特徴記述子 $D_3$ を用いて適合判定処理255を行い、明るさの観点で類似度が高くよく適合するビデオセグメントデータを絞り込み、最終結果256(図15における適合判定処理結果214と等価)を得る。

## 【0189】(2-3) 検索結果の提示とユーザフィードバックに基づく再検索 (S39~S41)

以上の検索ルールと適合判定処理に基づいて選択された $T$ 個の検索結果のビデオショットに関して、その情報をユーザインタフェース部229によりユーザ232に提示する(S39)。提示の方法は任意であるが、本実施の形態2では、例えば、最初にクエリーリストを提示したように、検索結果のビデオショットを代表するキーフレームデータやビデオショットの属性情報(ビデオショット内の被写体の名前など)を所定のルールに基づいて符号化して検索結果情報217としてクライアント側の検索処理部10Bに送信し、検索処理部10Bの検索結果受信・表示部228がサーバ側から送信されるキー

38

## 【0186】(e) 代表色に関する特徴記述子

クエリー $S_q$ 、テスト $S$ の間のマッチング評価尺度 $D(S_q, S)$ を下式で定める。ただし、 $[H_i^Y(S), H_i^U(S), H_i^V(S)]$ を $i$ 番目のサンプルのヒストグラム値とする。この $D(S_q, S)$ が小さいものほど類似度が高く、適合するものと判定する。

【0187】

10 【数23】

フレームデータや属性情報を符号化した検索結果情報217を受信して、ユーザインタフェース部229に表示可能な形式に変換してユーザインタフェース部229に引き渡す。以上の処理を経ることによって、ユーザ232は、検索結果をユーザインタフェース部229上に見ることができる。

【0190】次いで、ユーザフィードバックに基づく再検索の手順について説明する。ユーザは、最初はサーチプライオリティ241や信頼度242などに基づいて上記のように定められる検索ルールに従って得られた検索結果を確認するが、この結果がユーザの主観上、期待されるものと異なる場合が考えられる。そのような場合に柔軟に対処するため、本実施の形態2では、ユーザがサーチプライオリティを変更し、その変更値に基づいて再検索を行うことが可能となるようシステムを構成する

(S40)。

【0191】図20に、ユーザが特に着目したい主観的な意味と、特徴記述子との関連の一例を示す。同図からわかるように、特定の被写体が一定方向に移動するビデオショットをユーザがクエリーとして選択した際、検索結果として特に「動きの方向という特徴に対して類似するビデオショット」を期待していることが多いと考えることができる。しかし、最初の初期設定では、サーチプライオリティが信頼度に基づいて決定されるので、必ずしもこのような主観的な重み付けが検索過程に反映されていない可能性がある。

【0192】そこで、図20に示されるような主観的な意味をユーザが検索処理に反映できるよう、ユーザインタフェース部229等を介してサーチプライオリティ241を変更できるように構成する。例えば、上記の例で、ユーザ232は「動き方向の特徴記述子に関するサーチプライオリティをより高い値に設定する」ことを可能にする。このようにすることによって、検索処理は動き方向の特徴記述子の適合判定処理を先行して行うことになるので、最終的な検索結果に「動き方向」のファクタがより反映されやすくなる。



(21)

39

【0193】ユーザ232によって更新されたサーチプライオリティ241は、検索要求情報符号化・送信部227によって所定のルールで符号化され、検索要求情報216の一部として、ネットワークA225を介し、サーバ側の検索処理部10Aにおける検索用記述データ解析部211へ送信される。

【0194】検索用記述データ解析部211は、送られたサーチプライオリティ241の更新値を以降の検索処理に再利用できるよう、検索用記述データファイル210におけるビデオプログラム240に対応するサーチプライオリティ241の値を当該更新値によってアップデートする(S41)。

【0195】また、検索用記述データ解析部211は、一連の検索処理が終了した時点で、その時点で最新のサーチプライオリティ241の値を更新した状態で、検索用記述データファイル210を画像情報蓄積部5に再保存する。

【0196】このような構成をとることにより、ユーザは、同じビデオプログラムを再度同様の基準で検索する場合に、過去のユーザの主観要求を反映した形で再検索を行うようにすることができる。

【0197】また、複数の異なるユーザ232がシステムを利用するような場合、検索用記述データファイル210は、個々のユーザ232ごとにサーチプライオリティ241を格納可能なファイル構成とすることも考えられる。こうしておくことにより、適当なユーザ認証の手続きを組み込むことによって、個々のユーザ232に適した検索ルールを適宜再現することが可能となる。

【0198】以上のようにして更新されたサーチプライオリティ241を用いて、S37以降の検索処理を繰り返すことにより、再検索が行われることになる。これで図17に示す本システムにおける検索用記述データ生成および検索処理が終了する。

#### 【0199】(3) ビデオショット再生処理

以上述べた動作により、ユーザは所望の検索結果をユーザインタフェース部229上に引き出すことが可能となる。しかし、一般に検索結果は複数の検索候補を提示するものであり、それらをすべてビデオデータとして表示するにはサーバ側からクライアント側へ多量の情報伝送が必要となる。そこで、上述したように、ユーザインタフェース部229としては、ビデオショットを代表するキーフレームのようなデータを提示して、そこから選択的に映像再生が可能となるよう構成することにより、トータルの情報伝送量を効果的に削減することが可能である。そのようなシステム構成とするため、本実施の形態2では、図15および図16に示すビデオ再生サーバ218、ビデオ復号・再生部239、再生制御部235等を導入したものである。ユーザ232は、本機能を選択したビデオショットやビデオプログラムを再生するために利用する。

40

【0200】次に、ユーザ232による本機能を選択したビデオショットやビデオプログラムを再生動作について説明する。

【0201】まず、ユーザ232がユーザインタフェース部229上に提示されたキーフレームにより再生したいビデオショット244やビデオプログラム240を選択すると、その選択情報が検索結果再生要求233としてユーザインタフェース部229からビデオ再生制御部235へ出力される。

【0202】ビデオ再生制御部235では、所定の手順でその検索結果再生要求233を再生要求情報219に変換し、ネットワークA225を介してサーバ側のビデオ再生サーバ218の再生制御部220に送る。

【0203】再生制御部220では、再生要求情報219にコンテンツを特定するコンテンツ特定情報などを含めた再生コンテンツ特定情報221を生成するようにする。このような再生コンテンツ特定情報221は、もとも検索用記述データファイル210の中にコンテンツへのリンク情報として予め記述されており、ユーザに対し検索結果情報217を提示する際に、検索結果情報217にそのコンテンツ特定情報を付加してクライアント側に送信しておき、再生要求情報219として利用するという構成にすることもできる。

【0204】そして再生制御部220は、受信した再生要求情報219に基づく再生コンテンツ特定情報221を画像情報蓄積部5へ送り、画像情報蓄積部5から再生対象のビデオコンテンツデータ222を特定し、特定した再生対象のビデオコンテンツデータ222をビデオデータ送信部223へ送出する。

【0205】ビデオデータ送信部223では、再生対象のビデオコンテンツデータ222をリアルタイム配信可能なデータ形式の配信ビデオデータ224に変換した後、ネットワークB237へ送出する。

【0206】尚、ビデオデータ送信部223は、このような高速性を重視するネットワークB237に対してビデオデータを送出するにあたって、網内もしくはクライアント側で信頼性を確保することを可能とするためのデータ形式への変換を行うようにしても良い。例えば、ビデオデータをビデオのフレームを単位に分割し、その単位でタイムスタンプを付与したり、伝送単位のパケットに対する優先度の情報を付加したりすることなどが考えられる。

【0207】一方、クライアント側は、高速なネットワークB237を通じて送られてくる配信ビデオデータ224を、ビデオ復号・再生部239が受信して、別途ビデオ再生制御部235から渡される再生に必要な制御情報262に基づいて、ユーザインタフェース部229上にビデオ再生を行う。その際、ビデオ復号・再生部239およびビデオ再生制御部235は、伝送上の問題で伝送データの信頼性が低くなっているような場合は、適

(22)

41

宜、再生品質を補うための品質管理制御を行うようにする。

【0208】以上のように、本実施の形態2の画像検索システムおよび検索方法によれば、サーバ・クライアント型のシステム構成をとることによって、ローカルな環境だけでなく、IP網などの分散ネットワーク環境においてもビデオデータを効率的に検索・配信できる。

【0209】また、サーチプライオリティ241と信頼度242を検索用記述データとして画像情報蓄積部5に格納することにより、検索システムがそれらの値に基づいて共通的な検索ルールの決定指針を定めることができる。

【0210】また、サーチプライオリティ241の更新機能により、ユーザがビデオコンテンツの主観的な意味に基づいた優先付け検索を行うことができる。

【0211】さらに、特徴記述子の信頼度に基づく適合判定処理のスキップ制御により、検索処理速度を高速化することができ、システム効率を向上させることができる。

【0212】なお、上記実施の形態1、2では、扱うメディアはビデオデータのみとして述べてきたが、本発明では、これに限らず、別途、音声・オーディオデータなどが付加されたビデオデータについても同様の構成でシステムを構築することができる。また、オーディオデータに関して別途特徴記述子をベースとした検索系を用意してもよく、また、オーディオデータについては単にビデオデータに付随しており検索の対象にならないようなシステム構成も可能である。オーディオが付与される場合、ビデオショット再生処理については、ビデオデータに付随・同期するオーディオデータの再生もサポートするように構成すればよい。

【0213】また、本実施の形態2に述べたシステムは、特徴抽出部3および特徴記述子セット生成部4から構成され、特徴記述データを生成するモジュールと、画像情報蓄積部5から構成され、特徴記述データをデータベース登録するモジュールと、検索処理部10Aから構成され、ユーザからの検索要求を受け付けて検索を実行して結果を返す検索エンジンのモジュールと、主に検索処理部10Bから構成され、ユーザからの検索要求を検索エンジンに通知して返される検索結果をユーザに提示するクライアントモジュールについては、必ずしも同一ハードウェア・ソフトウェア上に構築される必要はなく、各々定められたインタフェースに従うことで独立にハードウェア、ソフトウェア実装することが可能である。サービスの観点からは、例えば、コンテンツプロバイダが特徴記述データを生成してデータベース登録を申請する。検索サービスプロバイダはそれらをデータベース登録し、検索処理の準備を整える。ユーザは、本実施の形態に述べたような、検索サービスプロバイダが指定する、もしくは標準化された規定に従って検索処理を実

42

行するクライアントツールで検索を実行する、という形態の運用方法が考えられる。

【0214】

【発明の効果】以上説明したように、本発明によれば、属性リストを参照してその属性情報に基づく検索を行なうと共に、特徴記述子に基づく検索を行うようにしたので、短時間かつ効率よく検索を行うことができる。

【0215】また、次の発明によれば、属性リストや特徴記述子をそれらのデータ構造を定義したシンタックスに基づいて生成したので、ネットワーク上に分散する複数の画像サーバに依存しない画像検索を行うことができる。

【0216】また、次の発明によれば、フレーム単位で特徴量を抽出すると共に、複数のフレームをまとめたビデオセグメント単位で特徴記述子を生成するようにしたので、動画画像の検索を容易に行うことができる。

【0217】また、次の発明によれば、入力画像データから画像特徴量を抽出し特徴記述子を生成すると共に、入力画像データの時空間構造に関連付けて特徴記述子を検索用記述データに記述して、入力画像データとともに蓄積し、ユーザからの検索要求に従って蓄積された入力画像データの検索用記述データを解析して特徴記述子を取り出し、取り出した特徴記述子に基づき適合判定処理を行って検索し、その結果をユーザに提示させるようにしたため、特徴記述子に基づく効率的な画像データの検索が可能となる。

【0218】さらに特徴記述子の信頼度を算出して、入力画像データの時空間構造に関連付けて特徴記述子および信頼度を検索用記述データに記述して、入力画像データとともに蓄積し、ユーザからの検索要求に従って蓄積された入力画像データの特徴記述子および信頼度を解析して取り出し、取り出した特徴記述子および信頼度に基づき適合判定処理を行って検索し、その結果をユーザに提示するようにしたため、特徴記述子および信頼度に基づく効率的な画像データの検索が可能となる。

【0219】また、適合判定処理の際、特徴記述子の信頼度に基づいて、特徴記述子による適合判定の必要性を評価し、適合性判定を行う必要がないと判断した場合、適合判定処理をスキップするようにしたので、検索処理速度を高速化することができ、検索効率をさらに向上させることができる。

【0220】また、特徴記述子の信頼度に基づいて各特徴記述子の検索使用順序を決定し、入力画像データの時空間構造に関連付けて特徴記述子、信頼度および検索使用順序を検索用記述データに記述し、ユーザからの検索要求に従って蓄積された入力画像データの特徴記述子、信頼度および検索使用順序を解析して取り出し、取り出した特徴記述子、信頼度および検索使用順序に基づき適合判定処理を行って検索し、その結果をユーザに提示するようにしたため、特徴記述子、信頼度および検索使

(23)

43

用順序に基づく効率的な画像データの検索が可能となる。

【0221】また、ユーザからの検索要求としての各特徴記述子の検索使用順序に従って、特徴記述子の検索順序を更新することにより、ユーザからの検索要求に見合う検索処理を実行することができる。

【0222】また、適合判定処理を行う際、各特徴記述子の信頼度に基づいて、検索使用順序で定められる各検索ステップにおける検索結果候補数を決定して適合判定処理を行うようにしたので、検索処理を高速化することが

【0223】さらに、入力画像データがビデオデータであり、検索によって特定された各ビデオデータを代表する各キー画像データを検索結果ユーザに提示し、提示された各キー画像のうちユーザによって特定のキー画像を選択すると、その選択されたキー画像を代表とするビデオデータを読み出して復号し再生するようにしたので、トータルの情報伝送量を効果的に削減することが可能である。

【図面の簡単な説明】

【図1】 本実施の形態1における画像検索システムの構成を説明するブロック図。

【図2】 画像データベース生成部1の処理を示すフローチャート。

【図3】 画像属性のテキスト情報からテキスト情報のインスタンス(XML文書)を生成する処理の一例を示す図。

【図4】 イントラ/インター適応符号化を示す図。

【図5】 特徴抽出処理および特徴記述子生成処理を含めたトータルの特徴記述子セット生成手順を示すフローチャート。

【図6】 図5に示すステップS15の特徴記述子セット生成処理の具体的な生成手順を示すフローチャート。

44

【図7】 動き分布を特徴づける量の一例を示す図。

【図8】 イントラフレームの記述子セットの意味を示す図。

【図9】 抽出された特徴量から特徴記述子セットを生成する処理の一例を示す図。

【図10】 特徴記述子セットを階層的に生成して記述する一例を示す図。

【図11】 画像検索部の処理を示すフローチャート。

【図12】 図11におけるステップS27の適合判定処理を示すフローチャート。

【図13】 画像データベース生成部の他の構成例を示す図。

【図14】 従来の画像検索処理システムの一例を示す図。

【図15】 本実施の形態2における画像検索システムのサーバ側の特徴部分の構成を示すブロック図。

【図16】 本実施の形態2における画像検索システムのクライアント側の特徴部分の構成を示すブロック図。

【図17】 本実施の形態2における画像検索システムの検索処理過程を示すフローチャート。

【図18】 本実施の形態2における画像検索システムが利用する検索用記述データファイルの構造を示す図。

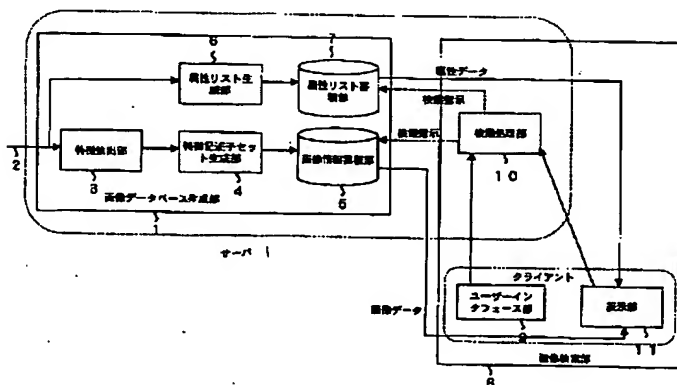
【図19】 適合判定処理部におけるサーチプライオリティに基づく特徴記述子の適合判定処理の一例を示す図。

【図20】 ユーザの主観的特徴と特徴記述子との関係を示す図。

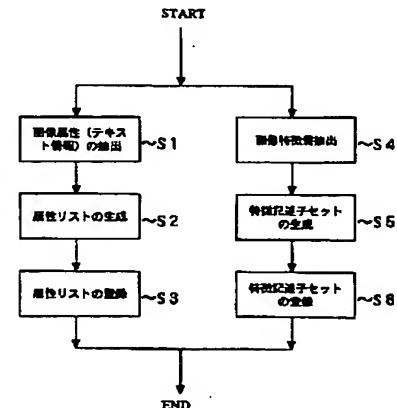
【符号の説明】

1 画像データベース生成部、3 特徴抽出部、4 特徴記述子セット生成部、5 画像情報蓄積部、6 属性リスト生成部、7 属性リスト蓄積部、8 画像検索部、9 ユーザインターフェース部、10 検索処理部、11 表示部、12 ビデオ圧縮符号化部。

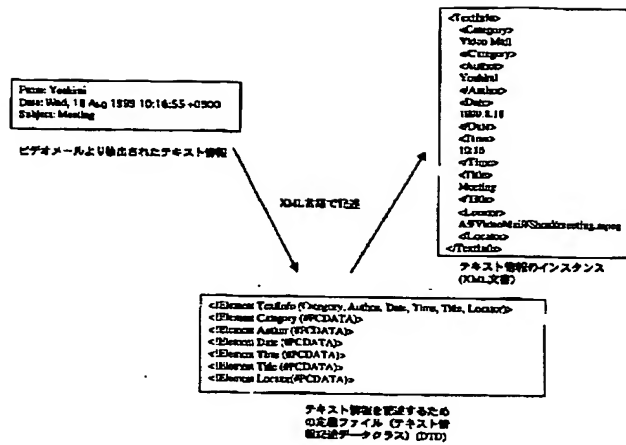
【図1】



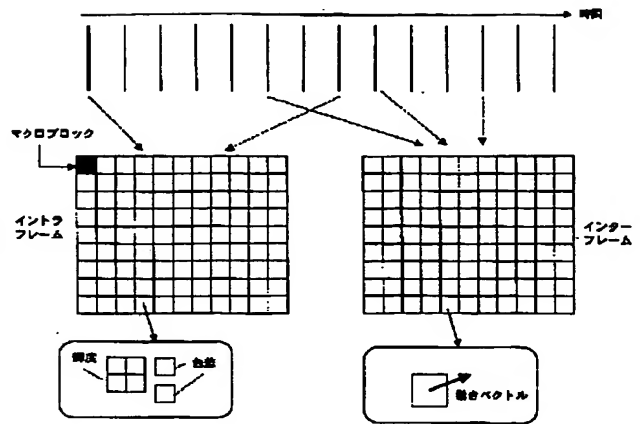
【図2】



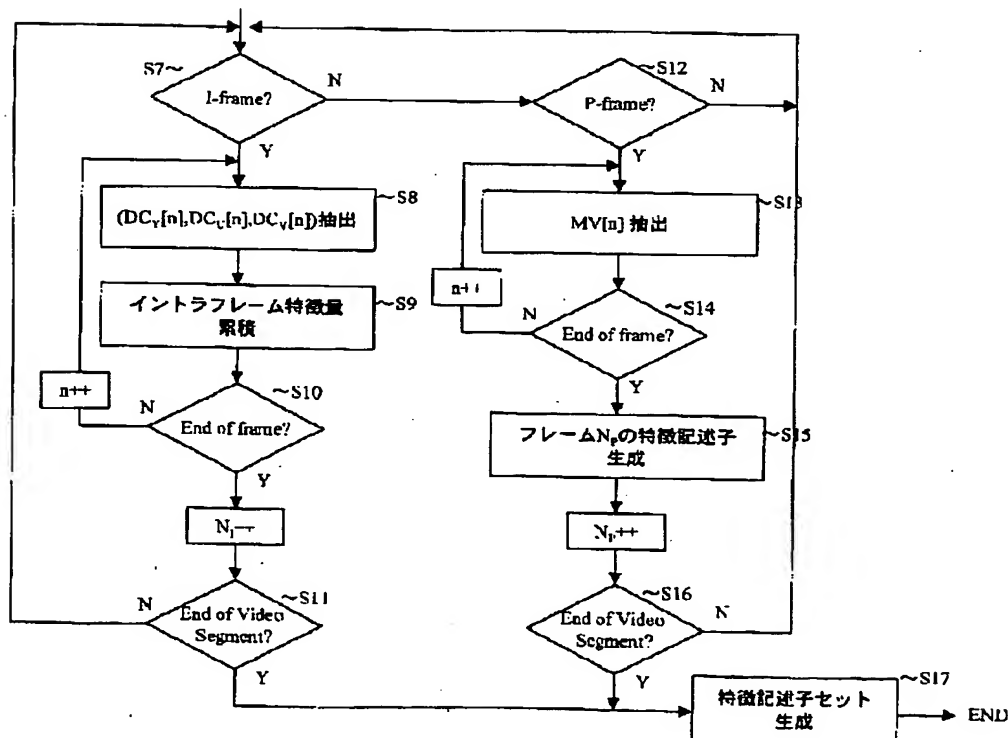
【図 3】



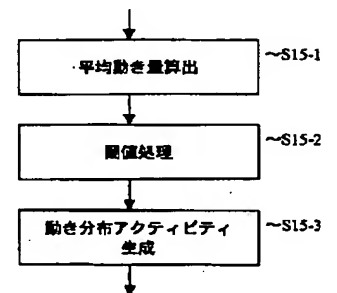
【図 4】



【図 5】

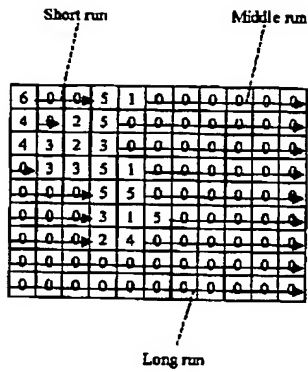


【図 6】

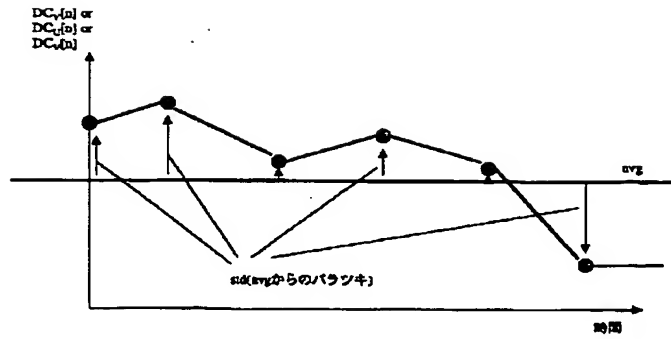


(25)

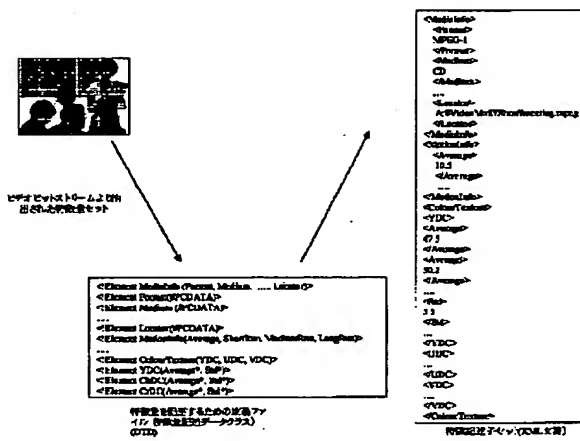
【図7】



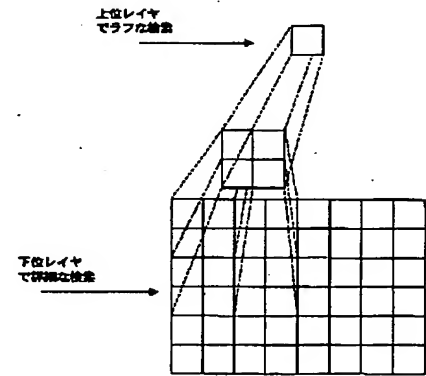
【図8】



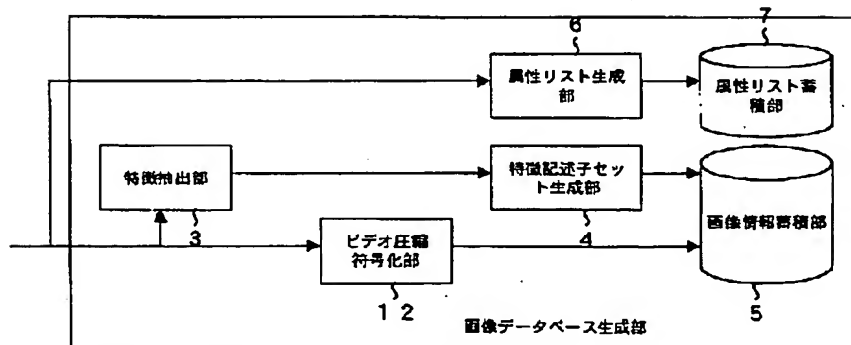
【図9】



【図10】

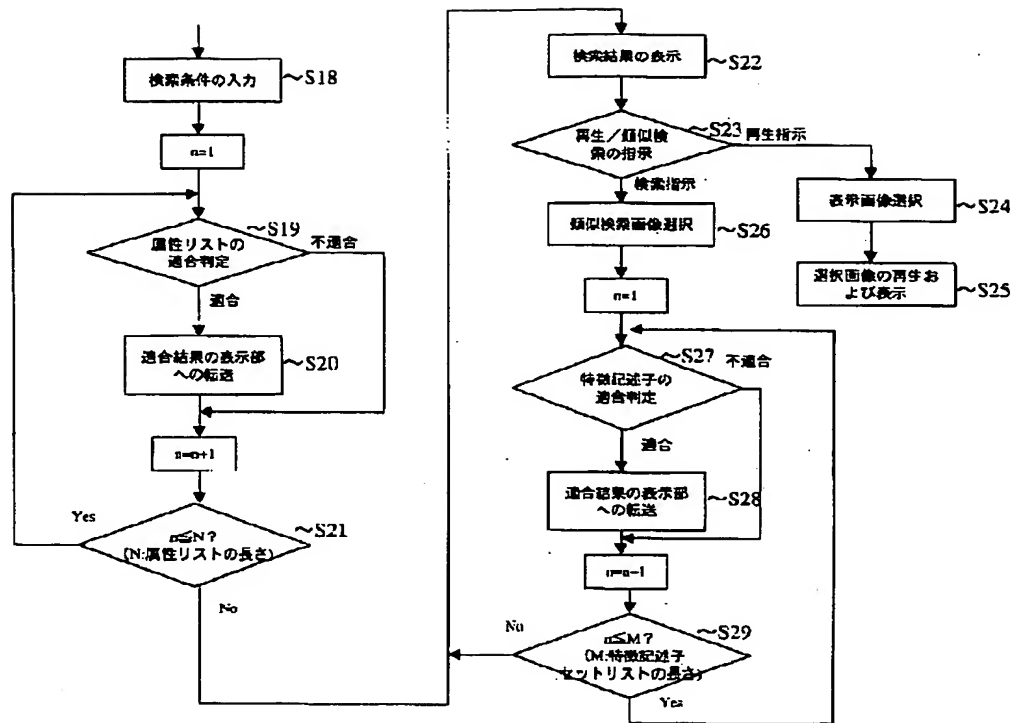


【図13】

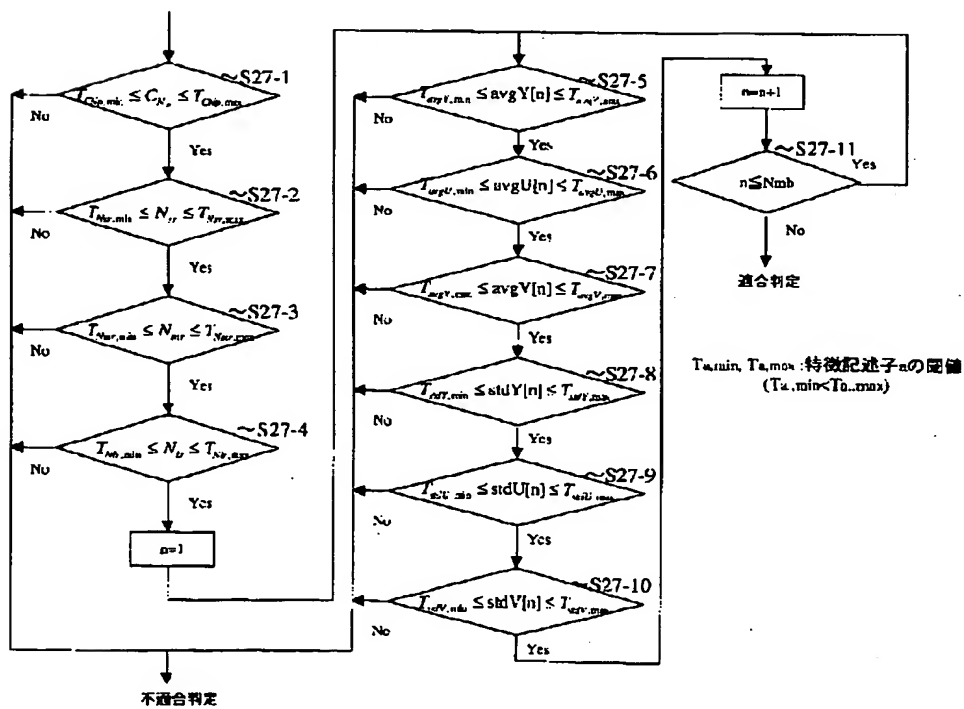


(26)

【図11】



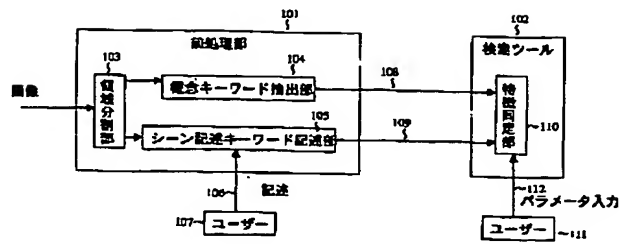
【図12】



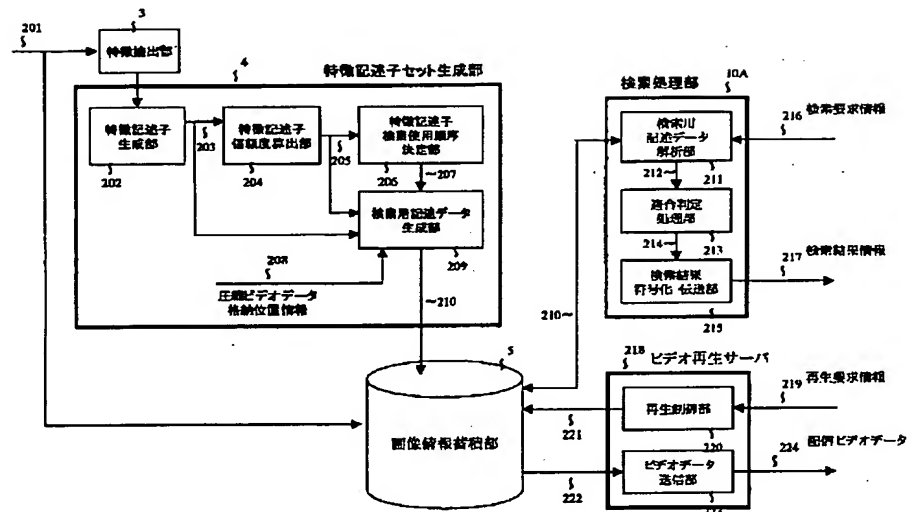


(27)

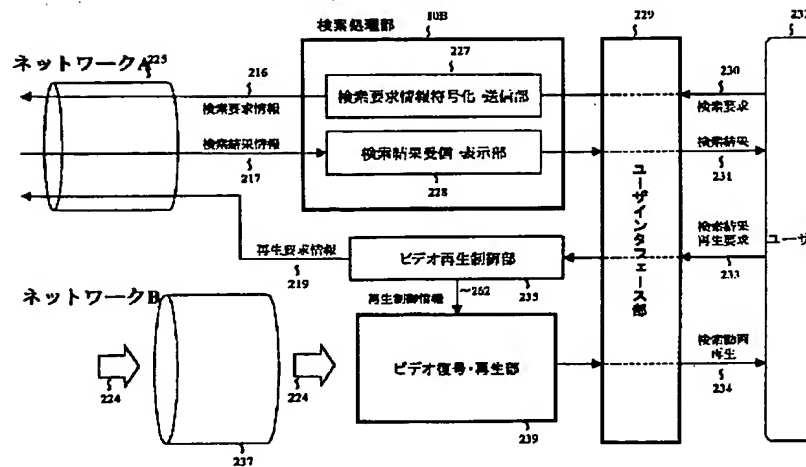
【図 14】



【図 15】

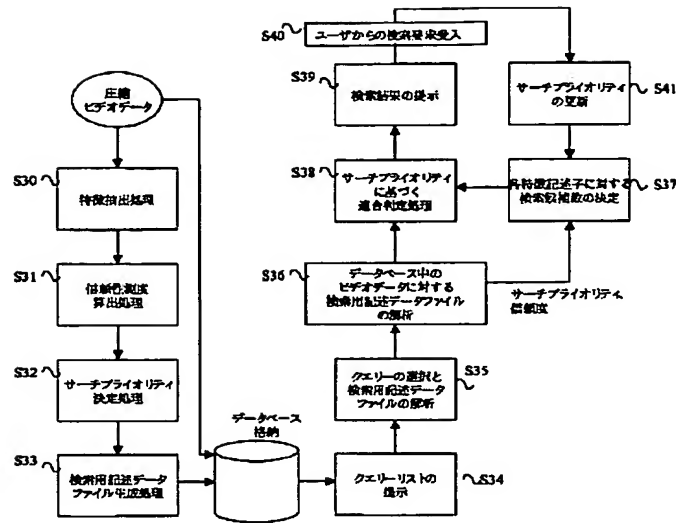


【図 16】

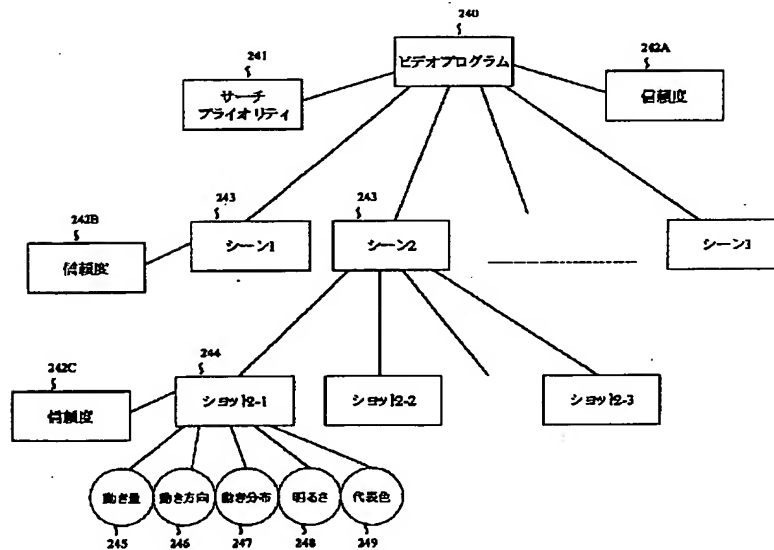


(28)

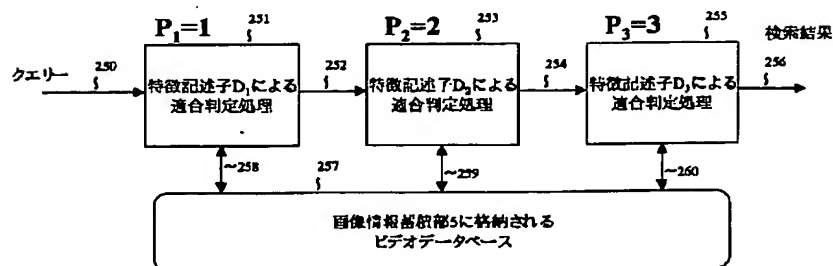
【図17】



【図18】



【図19】



(29)

【図20】

主観的意味	特徴記述子				
	動き量	動き方向	動き分布	明るさ	代表色
動き量の特徴が類似している	○				
動きの方向が類似している		○			
動きの西内分布の特徴が類似している			○		
映像中の平均的な明るさが類似している				○	
映像中の支配的な色情報が類似している					○

フロントページの続き

(72) 発明者 西川 博文  
 東京都千代田区丸の内二丁目2番3号 三  
 菱電機株式会社内

(72) 発明者 山田 悦久  
 東京都千代田区丸の内二丁目2番3号 三  
 菱電機株式会社内

(72) 発明者 浅井 光太郎  
 東京都千代田区丸の内二丁目2番3号 三  
 菱電機株式会社内

Fターム(参考) 5B050 BA10 EA04 GA08  
 5B075 ND06 ND12 NK06 NR16 PP12  
 PP22 PP23 PQ02  
 5L096 BA08 GA36 HA02  
 9A001 FF03 HH23 JJ01 JJ27